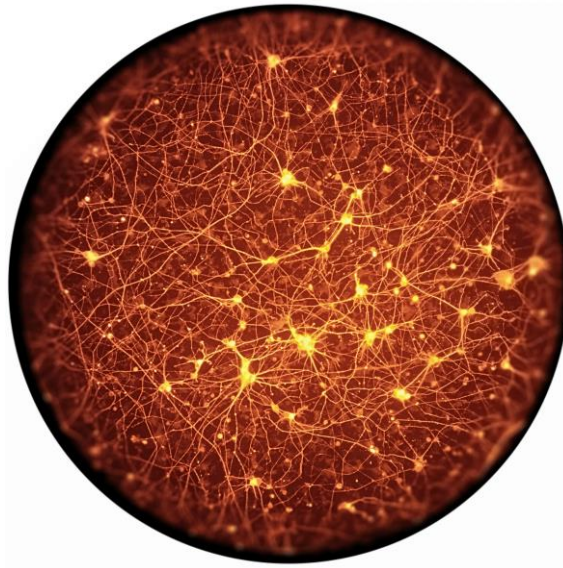# Development and characterization of the MoN-Light BoNT assay to determine the toxicity of botulinum neurotoxin in motor neurons differentiated from CRISPR-modified induced pluripotent stem cells

Brit-Maren Michaud Schjeide

DISSERTATION

to obtain the academic degree

"doctor rerum naturalium"

(Dr. rer. nat.)

in the scientific discipline "Biochemistry und Pathobiochemistry"

submitted to the

Faculty of Science

Institute of Nutritional Science – Department of Biochemistry of Nutrition

of the University of Potsdam

Berlin

Submitted: 21 April 2021

Defense: 18 August 2021

First supervisor: Prof. Dr. Gerhard Püschel
Second supervisor: Prof. Dr. Burkhard Kleuser

Thesis reviewers:
Prof. Dr. Gerhard Püschel (Universität Potsdam)
Prof. Dr. Burkhard Kleuser (Freie Universität, Berlin)
Prof. Dr. Ellen Fritsche (IUF Düsseldorf)

Table of Contents

d

e

Abbreviations

Full designation	Abbreviation
Degrees celcius	°C
delta	Δ
delta delta cycle threshold method	$2^{-\Delta\Delta Ct}$
Three prime	3'
Replacement, Reduction, and Refinement	3Rs
Five prime	5'
Five prime untranslated region	5'UTR
Adenine	A
Ascorbic acid	AA
Adeno-associated virus inverted terminal repeat	AAV2 ITR
Adeno-associated virus integration site 1 (T2)	AAVS1 (target 2)
Acetylcholine	ACh
Also known as	aka
Ampicillin resistance	AmpR
Annealing temperature	Ann. Temp
Analysis of variance	ANOVA
Action potential	AP
Anti-sense	AS
Adenosine triphosphatase	ATPase
Absorbance units	AU
Brain-derived neurotrophic factor	BDNF
Bovine growth hormone polyadenylation	bGH poly(A)
Botulinum neurotoxin	BoNT
Base pair	bp
Bovine serum albumin	BSA
Cytosine	C
Caenorhabditis elegans	C. elegans
Calcium, ion	Ca^{2+}
Calcium chloride	CaCl2
Catabolite activator protein binding site	CAP binding site
Chemokine (CC motif) receptor 5	CCR5
Complementary deoxyribonucleic acid	cDNA
Chromogranin A	CgA
Choline O-acetyltransferase	CHAT
Chromogranin A gene	*CHGA*
Selective small molecule GSK3 inhibitor	CHIR99021
CCAAT-enhancer-binding protein homologous protein	CHOP
Chromosome	Chr
Chromosome X	ChrX
Cytomegalovirus enhancer	CMV enhancer

Ciliary neurotrophic factor	CNTF
Carbon dioxide	CO_2
Clustered Regularly Interspaced Short Palindromic Repeats / CRISPR associated protein 9	CRISPR/Cas9
CRISPR RNA	crRNA
Chemiluminescent substrate (Roche)	CSPD
Cycle threshold	Ct
Carboxy-terminus	C-terminus
Day	D
4′,6-diamidino-2-phenylindole	DAPI
Inhibitor of the γ-secretase complex	DAPT
Dibutyryl cyclic adenosinemonophosphate	dbcAMP
Double control quantitative copy number polymerase chain reaction	dc-qcnPCR
Distilled water	dH_2O
Digitonin	DIG
Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12	DMEM-F12
Dorsomorphin homolog 1	DMH1
Dimethyl sulfoxide	DMSO
Deoxyribonucleotide triphosphate	dNTP
Double strand break	DSB
Dithiothreitol	DTT
Escherichia coli	E.coli
Ethylene diamine tetraacetic acid	EDTA
exempli gratia	e.g.
Elongation factor alpha	Ef1α
Elongation factor - human T-cell leukemia virus	Ef1-HTLV
Ethylene-bis(oxyethylenenitrilo)tetraacetic acid	EGTA
Enzyme-linked Immunosorbent Assay	ELISA
Enhanced specificity Cas9	eSpCas9
et alia	*et al*
Ethidium bromide	EtBr
Ethanol	EtOH
Degrees of freedom	f
Forward	F/Fwd
Fast alkaline phosphatase	Fast AP
Fetal calf serum	FCS
FIJI is just Image J	Fiji
Guanine	G
Glyceraldehyde 3-phosphate dehydrogenase	GAPDH
GATA binding protein 1	GATA1
Cytosine / guanine content	GC content
(Genomic) deoxyribonucleic acid	(g)DNA
Glial cell-derived neurotrophic factor	GDNF

Green fluorescent protein	GFP
Gaussia luciferase	GLuc
Golgi matrix protein of 130 kDa	GM130
Gene of interest	goi
Golgin-97, RanBP2α, Imh1p and trans golgi p230	GRIP
Genomic safe harbor	GSH
Hour	h
Water	H_2O
Homology arm	HA
Haplotype map	HapMap
Motor neuron and pancreas homeobox 1	HB9, MNX1
Heavy chain	HC
Heavy chain C-terminus	H_C
Hydrogen chloride	HCl
Homology-directed recombination	HDR
Human embryonic kidney	HEK
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid	HEPES
(Human induced) pluripotent stem cells	(hi)PSCs
Human Immunodeficiency Virus	HIV
Heavy chain N-terminus	H_N
Helix pomatia agglutinin	HPA
(Human) proopiomelanocortin	(h)POMC
Hypoxanthine Phosphoribosyltransferase 1	HPRT1
Immunofluorescence	IF
iPS cell line clone 4, derived from IMR90-4 human fetal lung fibroblasts	IMR90-4
Islet1	ISL1
Potassium, ion	K^+
Potassium HEPES buffered saline	K^+-HBS
(Kilo) base pairs	(k)bp
Potassium chloride	KCl
Lysogeny broth	LB
Light chain	LC
Lethal dose 50	LD_{50}
Large dense core vesicle	LDCV
LIN-11, Isl-1 and MEC-3	LIM Domain
Lin-28 homolog A	LIN28A
Luminescence units	LU
Molar	M
Mouse bioassay	MBA
Multiple cloning site	MCS
Magnesium, ion	Mg^{2+}
Magnesium sulfate	$MgSO_4$

Minute	min
Motor neuron	MN
Motor neuron light botulinum neurotoxin assay	MoN-Light BoNT
(Messenger) ribonucleic acid	(m)RNA
Mammalian uncoordinated-18	Munc18
Number	n
Sodium, ion	Na^+
Sodium HEPES buffered saline	Na^+-HBS
Sodium chloride	NaCl
Nanog homeobox	NANOG
Sodium hydroxide	NaOH
New England Biolabs	NEB
Neurogenin 2	NEUROG2
Non-homologous end joining	NHEJ
homeodomain transcription factor NK6 homeobox 1	NKX6.1
Nuclear lysis buffer	NL buffer
Neuropeptide-like protein	NLP-21
Nonionic polyoxyethylene 40	NP-40
N-ethylmaleimide sensitive fusion protein	NSF
No tag	nt
Amino terminus	N-terminus
Octamer -binding transcription factor 4	OCT4
Oligodendrocyte transcription factor 2	OLIG2
Oligo-deoxythymidine	Oligo (dt)
Overnight	ON
Open reading frame	ORF
Origin of replication	ori
Probability (-value)	p(-value)
Pellets 1 through 5	P1-5
Plasmid AAVS1-Puromycin-Multiple cloning site	pAAVS1-P-MCS
Penicillin / Streptomycin	P/S
Protospacer adjacent motif	PAM
Paired box 6	PAX6
Phosphate-buffered saline	PBS
Phosphate-buffered saline – Triton X	PBS-T
Pheochromocytoma 12 cells	PC12 cells
Polymerase chain reaction	PCR
Paraformaldehyde	PFA
potentia Hydrogenii	pH
Protease inhibitor cocktail	PI cocktail
Poly-L-lysine	PLL
Purmorphamine	PMA
Progenitor motor neuron	pMN

Polynucleotide Kinase	PNK
Peptidylprolyl isomerase A	PPIA
Protein phosphatase 1 regulatory subunit 12C	PPP1R12C
Puromycin	puro
Puromycin resistance	PuroR
Quantitative polymerase chain reaction	qPCR
Reverse	R
Retinoic Acid	RA
Restriction enzyme	RE
Random insertion	RI
RB binding protein 7	RBBP7
Ribonucleic acid	RNA
Reference gene	ref gene
Reverse Oriented Splice Acceptor, Clone 26)	ROSA26
Rotations per minute	rpm
Roswell Park Memorial Institute 1640	RPMI1640
Ribosomal Protein S23	RPS23
Room temperature	RT
Reverse-transcription polymerase chain reaction	RT-PCR
Sense	S
Synthesis/Gap2 phase	S/G_2 phase
Supernatants 1 through 5	S1-5
Thiol group bond	SS
Smoothened agonist	SAG
4-(imidazol-2-yl)benzamide	SB431542
Santa cruz	SC
Secretogranin II gene	*SCG2*
Standard deviation	SD
Sodium-Dodecyl-Sulfate	SDS
Second	sec
Sequence	seq
Subcellular fractionation buffer	SF buffer
Secretogranin II	SgII
Single guide RNA	sgRNA
Cell line established from the adrenal tumor tissue resected after treatment from a 20-month-old boy with neuroblastoma	SIMA
Synaptosomal-associated protein, 25kDa	SNAP-25
SNAP Receptor	SNARE
Single nucleotide polymorphism	SNP
Super Optimal broth	SOB
Super Optimal broth with Catabolite repression	SOC
SRY-box transcription factor 2	SOX2
Saline Sodium Citrate	SSC

Synaptic vesicle	SV
Synaptic vesicle protein 2	SV2
Synaptophysin	Syp
Synaptotagmin	Syt
Thymine	T
Tris-acetate-EDTA (TAE) buffer	TAE
Transcription activator-like effector nucleases	TALENS
Temperature	Temp
Tobacco etch virus	TEV
Trans-golgi network	TGN
Trans-golgi network protein 46kDa	TGN46
Hybridization temperature	T_{hyb}
Melting temperature	T_m
Thymosin Beta 15B	TMSB15B
Transactivating crRNA	tracrRNA
Unit	U
University of California, Santa Cruz	UCSC
Volt	V
Vesicular acetylcholine transporter	VAChT
Synaptobrevin	VAMP
Relative centrifugal force / times gravity	x g
Yellow fluorescent protein	YFP
Wildtype	WT
Zinc finger nuclease	ZFN

List of Figures

viii

List of Tables

Abstract

Botulinum neurotoxin (BoNT) is produced by the anaerobic bacterium *Clostridium botulinum*. It is one of the most potent toxins found in nature and can enter motor neurons (MN) to cleave proteins necessary for neurotransmission, resulting in flaccid paralysis. The toxin has applications in both traditional and esthetic medicine. Since BoNT activity varies between batches despite identical protein concentrations, the activity of each lot must be assessed. The gold standard method is the mouse lethality assay, in which mice are injected with a BoNT dilution series to determine the dose at which half of the animals suffer death from peripheral asphyxia. Ethical concerns surrounding the use of animals in toxicity testing necessitate the creation of alternative model systems to measure the potency of BoNT.

Prerequisites of a successful model are that it is human specific; it monitors the complete toxic pathway of BoNT; and it is highly sensitive, at least in the range of the mouse lethality assay. One model system was developed by our group, in which human SIMA neuroblastoma cells were genetically modified to express a reporter protein (GLuc), which is packaged into neurosecretory vesicles, and which, upon cellular depolarization, can be released – or inhibited by BoNT – simultaneously with neurotransmitters. This assay has great potential, but includes the inherent disadvantages that the GLuc sequence was randomly inserted into the genome and the tumor cells only have limited sensitivity and specificity to BoNT. This project aims to improve these deficits, whereby induced pluripotent stem cells (iPSCs) were genetically modified by the CRISPR/Cas9 method to insert the GLuc sequence into the AAVS1 genomic safe harbor locus, precluding genetic disruption through non-specific integrations. Furthermore, GLuc was modified to associate with signal peptides that direct to the lumen of both large dense core vesicles (LDCV), which transport neuropeptides, and synaptic vesicles (SV), which package neurotransmitters. Finally, the modified iPSCs were differentiated into motor neurons (MNs), the true physiological target of BoNT, and hypothetically the most sensitive and specific cells available for the MoN-Light BoNT assay.

iPSCs were transfected to incorporate one of three constructs to direct GLuc into LDCVs, one construct to direct GLuc into SVs, and one "no tag" GLuc control construct. The LDCV constructs fused GLuc with the signal peptides for proopiomelanocortin (hPOMC-GLuc), chromogranin-A (CgA-GLuc), and secretogranin II (SgII-GLuc), which are all proteins found in the LDCV lumen. The SV construct comprises a VAMP2-GLuc fusion sequence, exploiting the SV membrane-associated protein synaptobrevin (VAMP2). The no tag GLuc

expresses GLuc non-specifically throughout the cell and was created to compare the localization of vesicle-directed GLuc.

The clones were characterized to ensure that the GLuc sequence was only incorporated into the AAVS1 safe harbor locus and that the signal peptides directed GLuc to the correct vesicles. The accurate insertion of GLuc was confirmed by PCR with primers flanking the AAVS1 safe harbor locus, capable of simultaneously amplifying wildtype and modified alleles. The PCR amplicons, along with an insert-specific amplicon from candidate clones were Sanger sequenced to confirm the correct genomic region and sequence of the inserted DNA. Off-target integrations were analyzed with the newly developed dc-qcnPCR method, whereby the insert DNA was quantified by qPCR against autosomal and sex-chromosome encoded genes. While the majority of clones had off-target inserts, at least one on-target clone was identified for each construct.

Finally, immunofluorescence was utilized to localize GLuc in the selected clones. In iPSCs, the vesicle-directed GLuc should travel through the Golgi apparatus along the neurosecretory pathway, while the no tag GLuc should not follow this pathway. Initial analyses excluded the CgA-GLuc and SgII-GLuc clones due to poor quality protein visualization. The colocalization of GLuc with the Golgi was analyzed by confocal microscopy and quantified. GLuc was strongly colocalized with the Golgi in the hPOMC-GLuc clone ($r = 0.85\pm0.09$), moderately in the VAMP2-GLuc clone ($r = 0.65\pm0.01$), and, as expected, only weakly in the no tag GLuc clone ($r = 0.44\pm0.10$). Confocal microscopy of differentiated MNs was used to analyze the colocalization of GLuc with proteins associated with LDCVs and SVs, SgII in the hPOMC-GLuc clone ($r = 0.85\pm0.08$) and synaptophysin in the VAMP2-GLuc clone ($r = 0.65\pm0.07$). GLuc was also expressed in the same cells as the MN-associated protein, Islet1.

A significant portion of GLuc was found in the correct cell type and compartment. However, in the MoN-Light BoNT assay, the hPOMC-GLuc clone could not be provoked to reliably release GLuc upon cellular depolarization. The depolarization protocol for hPOMC-GLuc must be further optimized to produce reliable and specific release of GLuc upon exposure to a stimulus. On the other hand, the VAMP2-GLuc clone could be provoked to release GLuc upon exposure to the muscarinic and nicotinic agonist carbachol. Furthermore, upon simultaneous exposure to the calcium chelator EGTA, the carbachol-provoked release of GLuc could be significantly repressed, indicating the detection of GLuc was likely associated with vesicular fusion at the presynaptic terminal. The application of the VAMP2-GLuc clone in the MoN-Light BoNT assay must still be verified, but the results thus far indicate that this clone could be appropriate for the application of BoNT toxicity assessment.

Zusammenfassung

Botulinum neurotoxin (BoNT) wird von dem obligat anaeroben Bakterium *Clostridium botulinum* produziert. Es ist eines der giftigsten natürlich vorkommenden Toxine. Nach Aufnahme in den Körper dringt es in Motorneurone ein und spaltet spezifische Proteine, die für die Freisetzung des Neurotransmitters Acetylcholin notwendig sind. Dadurch kommt es zu einer schlaffen Lähmung der Muskulatur, die zu einer peripheren Asphyxie führt. Trotz seiner hohen Toxizität wird BoNT als Therapeutikum in der klassischen und kosmetischen Medizin genutzt. Da die Aktivität des biosynthetisch gewonnenen Toxins zwischen einzelnen Chargen trotz gleicher Proteinkonzentration stark variiert, muss die Aktivität jeder Präparation getestet werden. Dafür ist der Goldstandard der Mausletalitäts-Test, bei dem den Tieren unterschiedliche Dosen des Toxins injiziert werden und die Dosis ermittelt wird, bei der die Hälfte der Tiere verstirbt. Wegen der damit verbundenen ethischen Probleme wird dringend nach Ersatzverfahren für diesen Tierversuch gesucht.

Ein Ersatzverfahren muss folgende Bedingungen erfüllen: Es muss humanspezifisch sein; es muss alle Teilschritte der BoNT-Wirkung messen; und es muss eine hohe Empfindlichkeit haben, die in der gleichen Größenordnung wie der Maus-Letalitätstest liegt. Es wurde bereits ein Testsystem von unserer Gruppe entwickelt, bei dem humane SIMA-Neuroblastomzellen genetisch so modifiziert wurden, dass sie ein Reporterprotein (GLuc) exprimieren. Dieses wurde in neurosekretorische Vesikel verpackt und durch Depolarisation der Zellen gleichzeitig mit Neurotransmittern freigesetzt. Die Freisetzung wurde durch BoNT gehemmt. Obwohl dieser Assay großes Potential hat, wird seine Anwendbarkeit durch inhärente Nachteile eingeschränkt, da die GLuc-Sequenz zufällig in das Genom eingefügt wurde und die Tumorzellen nur eine begrenzte Sensitivität und Spezifität gegenüber BoNT haben. Diese Dissertation hatte zum Ziel, diese Defizite zu verbessern. Zu diesem Zweck wurden induzierte pluripotente Stammzellen (iPSCs) durch die CRISPR/Cas9-Methode genetisch modifiziert, um die GLuc-Sequenz in den genomischen Safe-Harbor-Lokus AAVS1 einzufügen, wodurch ausgeschlossen wird, dass durch unspezifische Integrationen ins Genom die Funktion anderer Gene gestört wird. Darüber hinaus wurde GLuc so modifiziert, dass sie mit Signalpeptiden versehen wurde, die sie zum Lumen sowohl von „Large Dense Core" Vesikeln (LDCV), die Neuropeptide transportieren, als auch von synaptischen Vesikeln (SV), die Neurotransmitter verpacken, führen. Schließlich wurden die modifizierten iPSCs in Motorneurone (MNs) differenziert, der eigentlichen physiologischen Zielstruktur von BoNT,

die mutmaßlich am empfindlichsten und spezifischsten auf BoNT reagieren und daher für den MoN-Light BoNT-Assay am geeignetsten ein sollten.

iPSCs wurden transfiziert, um eines von drei Konstrukten zu integrieren. 1) ein Konstrukt, das GLuc in LDCVs leitet, 2) ein Konstrukt, das GLuc durch Fusion mit VAMP2 in SVs leitet und 3) ein "no tag" GLuc-Kontrollkonstrukt. Die LDCV-Konstrukte enthielten die Signalpeptide Proopiomelanocortin (hPOMC), Chromogranin-A (CgA) und Secretogranin II (SgII). Die VAMP2-GLuc-Fusion transportiert GLuc in SVs, so dass Neurotransmitter und GLuc gemeinsam und nicht, wie bei den anderen Konstrukten parallel, aus unterschiedlichen Vesikeln freigesetzt werden. Die "no tag GLuc"-Kontrolle wurde erstellt, um die Lokalisation von GLuc, die ohne Sortierungssignal in der Zelle exprimiert wird, mit der GLuc mit Sortierungssignalen für die unterschiedlichen Vesikel zu vergleichen.

Die Klone wurden charakterisiert, um sicherzustellen, dass die GLuc-Sequenz ausschließlich in den AAVS1-Safe-Harbor-Lokus eingebaut wurde und dass die Signalpeptide GLuc zu den richtigen Vesikeln leiten. Die korrekte Insertion von GLuc wurde durch PCR mit Primern bestätigt, die den AAVS1-Lokus flankieren und in der Lage sind, gleichzeitig Wildtyp- und modifizierte Allele zu amplifizieren. Die PCR-Amplikons wurden zusammen mit einem Insert-spezifischen Amplikon von Kandidatenklonen mittels Sanger-Sequenzierung untersucht, um die korrekte genomische Region und Sequenz der eingefügten DNA zu bestätigen. Mögliche Integrationen außerhalb der Zielregion wurden mit der neu entwickelten dc-qcnPCR analysiert, wobei die Insert-DNA mittels qPCR gegen autosomal und geschlechts-chromosomal kodierte Gene quantifiziert wurde. Auch wenn die Mehrzahl der analysierten Klone Off-Target-Integrationen enthielt, konnte für jedes Konstrukt mindestens ein vollständig On-Target-homozygoter Klon identifiziert werden.

Schließlich wurden die GLuc in ausgewählten Klonen durch Immunfluoreszenz lokalisiert. In iPSCs sollte die GLuc mit Sortierungssequenzen für Vesikel durch den Golgi-Apparat entlang des neurosekretorischen Weges wandern, während die „no tag" GLuc diesem Weg nicht folgen sollte. Anfängliche Analysen schlossen die CgA-GLuc- und SgII-GLuc-Klone aufgrund der schlechten Qualität der Proteinvisualisierung aus. Die Kolokalisation von GLuc mit dem Golgi-Apparat wurde mittels konfokaler Mikroskopie analysiert und quantifiziert. GLuc war im hPOMC-GLuc-Klon sehr stark ($r = 0,85\pm0,09$), im VAMP2-GLuc-Klon mäßig ($r = 0,65\pm0,01$) und im no tag GLuc-Klon erwartungsgemäß nur schwach ($r = 0,44\pm0,10$) mit Golgi-Markern assoziiert. Nach der Differenzierung in MNs wurde die Koexpression von GLuc mit dem MN-assoziierten Protein Islet1 bestätigt. Konfokale

Mikroskopie von MNs wurde angewandt, um die Kolokalisation von GLuc mit Proteinen zu quantifizieren, die mit LDCVs und SVs assoziiert sind, nämlich SgII mit der hPOMC-GLuc (r = 0,85±0,08) und Synaptophysin mit VAMP2-GLuc (r = 0,65±0,07).

Ein signifikanter Anteil von GLuc wurde im richtigen Zelltyp und Kompartiment gefunden. Im MoN-Light BoNT-Assay wurde die GLuc jedoch nicht zuverlässig durch Depolarisation aus dem hPOMC-GLuc-Klon freigesetzt. Das für die SIMA-hPOMC-Gluc-Zellen entwickelte Depolarisationsprotokoll muss für hPOMC-GLuc weiter optimiert werden, um eine zuverlässige und spezifische Freisetzung von GLuc bei Exposition gegenüber einem Stimulus zu erreichen. Andererseits konnte die GLuc aus dem VAMP2-GLuc-Klon durch Stimulation mit dem muskarinischen und nikotinischen Agonisten Carbachol freigesetzt werden. Die Carbachol-abhängige Freisetzung der GLuc konnte mit dem Calcium-Chelator EGTA unterdrückt werden, was darauf hindeutet, dass die Freisetzung der GLuc wahrscheinlich von der Fusion synaptischer Vesikel am präsynaptischen Terminal abhängig ist. Die Anwendung des VAMP2-GLuc-Klons im MoN-Light BoNT-Assay muss noch verifiziert werden, aber die bisherigen Ergebnisse deuten darauf hin, dass dieser Klon für die Anwendung der BoNT-Toxizitätsbewertung geeignet sein könnte.

1 Introduction

Botulinum Neurotoxins (BoNT) are some of the most toxic naturally occurring substances on Earth. Exposure to BoNT disrupts the protein complex necessary for synaptic vesicle fusion with the presynaptic membrane of the motor neuron end junction, inhibiting neurotransmitter release into the synaptic cleft. The subsequent blockade of signal propagation causes flaccid muscle paralysis. Despite the inherent danger in overexposure to the toxin, BoNT has found its place in treatment of both cosmetic and medical conditions, appropriate for a wide variety of spastic muscle disorders and as pain alleviators. Their implementation has proven to be an important factor in the improvement in the quality of life and condition of patients receiving the toxin as medication. Since this substance has a narrow therapeutic index and potentially lethal side effects, the efficacy of these treatments has been tested, developed, and quality controlled with the help of animal testing, whereby hundreds of thousands of mice are sacrificed each year to determine the toxicity of each pharmaceutical preparation[1].

As technology improves and the ethics of animal testing are more rigorously questioned, the opportunity to develop safe and reliable alternatives to animal testing is evident. The aim of this project is to improve and specialize an *in vitro* cell-based toxicity assay to measure the potency of individual pharmaceutical charges of BoNT. A prototype of this experimental method, termed Random_Insertion-hPOMC1-26GLuc, was developed in SIMA neuroblastoma cells, in which a reporter gene sequence (Gaussia luciferase, GLuc) was stably transfected into the SIMA genomic DNA (gDNA). The GLuc, which was randomly inserted into the SIMA genome, contains a signal sequence to sort the reporter protein into neurosecretory vesicles. Upon stimulation of the nerve cells, both neurotransmitters and the reporter protein in the neurosecretory vesicles would be released into the medium surrounding the cells, which can be collected and analyzed. Upon exposure of BoNT, neurosecretory vesicle fusion to the plasma membrane is blocked and neurotransmitter and reporter protein release should be inhibited in parallel. Preliminary experiments indicated that this method had great potential: the sensitivity of the cells to the BoNT was similar to that in the mouse lethality test and the method was simple and reproducible[2]. In order to improve the method, this project involved making the following modifications to the assay:

1. The cells are genetically modified with CRISPR/Cas9, specifically inserting the reporter gene at the AAVS1 safe harbor site in order to avoid any off-target integration sites and therefore potentially harmful effects on the cells.

2. The modified cells are human induced pluripotent stem cells (hiPSCs), which retain the ability to be differentiated into any cell type.

3. Post-modification, the cells are differentiated into motor neurons, the physiological target of BoNTs.

A summary of the assay, designated the MoN-Light BoNT assay, is visualized in Figure 1.
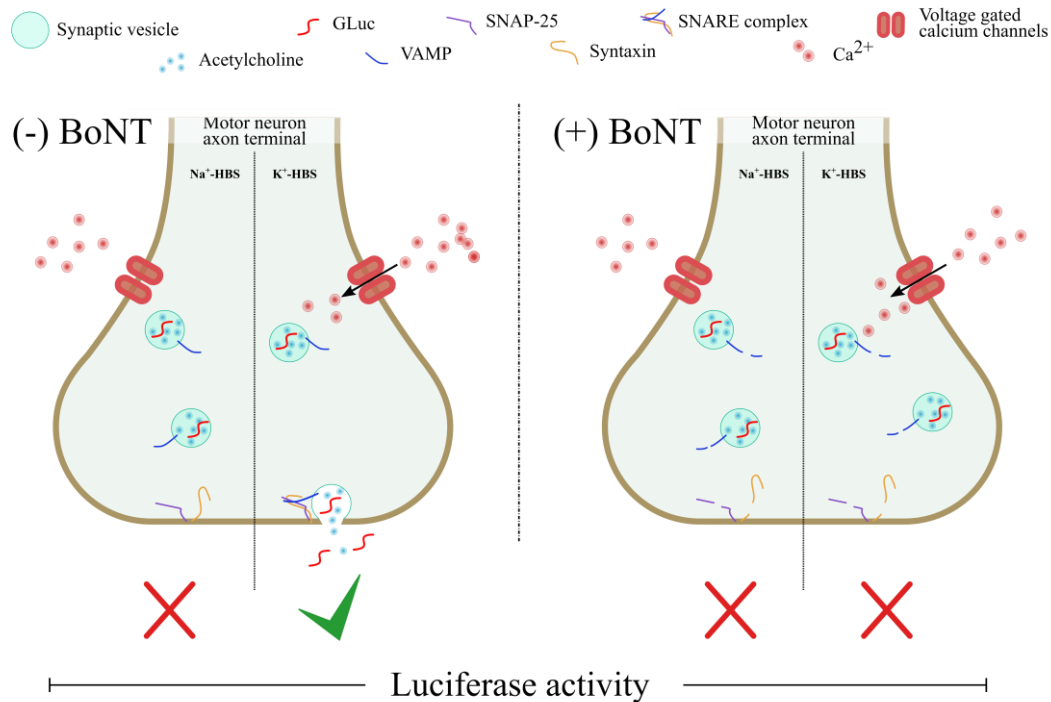
Figure 1: Visual representation of MoN-Light BoNT assay. Under normal conditions ((-) BoNT), exposure to the control buffer (Na^+-HBS) does not initiate Ca^{2+} influx and luciferase is not released into the solution surrounding the cell. Exposure to the depolarization buffer (K^+-HBS) causes an influx of Ca^{2+} into the presynaptic terminal, initiating the creation of the SNARE complex, and allowing the release of neurotransmitters and luciferase into the surrounding solution, resulting in detection of increased luciferase activity. Upon exposure to a high concentration of BoNT ((+) BoNT), the members of the SNARE complex are cleaved (depending on BoNT serotype). The circumstances under the control buffer remain the same. Exposure to the depolarization buffer still instigates Ca^{2+} influx, however the SNARE complex is disrupted and the vesicle cannot fuse to the presynaptic membrane, blocking neurotransmitter and luciferase release. No increase in luciferase activity is detected.

To elucidate the intricacies of this project, the following passages will provide an overview of the development of the nervous system, including motor neurons and neurosecretory vesicles. The role of the SNARE protein complex in neurotransmission will be clarified. Next BoNT will be described, including its protein structure and molecular mechanism. Furthermore, the current methods to assess BoNT toxicity will be summarized. Finally, the advent of induced pluripotent stem cells and the advances of technologies to genetically modify organisms and cells will be addressed.

Section 1.1

1.1 Development of the nervous system

1.1.1 Neuronal development

The differentiation of iPSCs into motor neurons emulates the processes taking place in early embryogenesis. Early in embryogenesis a process called gastrulation leads to the formation of the three primary germ layers, the endoderm, the mesoderm, and the ectoderm. A cylinder of mesodermal cells begins to extend through the embryo, forming the embryonic midline called the notochord. Directly above the notochord is a specific part of the ectoderm called the neuroectoderm, from which the nervous system develops. The notochord is responsible for sending the appropriate signals to induce neural differentiation: the neuroectodermal cells differentiate into neural precursor cells in a process called neurulation, in which the neural plate develops along the midline ectoderm. The neural plate begins to fold into the neural tube, which subsequently forms the brain and spinal cord (process summarized in Figure 2). Within the neural tube there are neural precursor cells, stem cells that become neuroblasts, which then differentiate into neurons[3]. Specifically, within the ventral neural tube multiple classes of neurons begin to develop, including both upper and lower motor neurons (MNs). Neuronal fate is strictly linked to cellular position and the extrinsic signals reaching these locations[4]. The highly organized spatial and temporal coordination of endogenous signaling molecules regulates gene expression at very specific timepoints of fetal development. An example of this inductive molecular signaling molecules is retinoic acid (RA), which is secreted by the somites and targets the gene expression through modulation of transcription factors, and whose dysregulation can severely disrupt neural development[5,6].
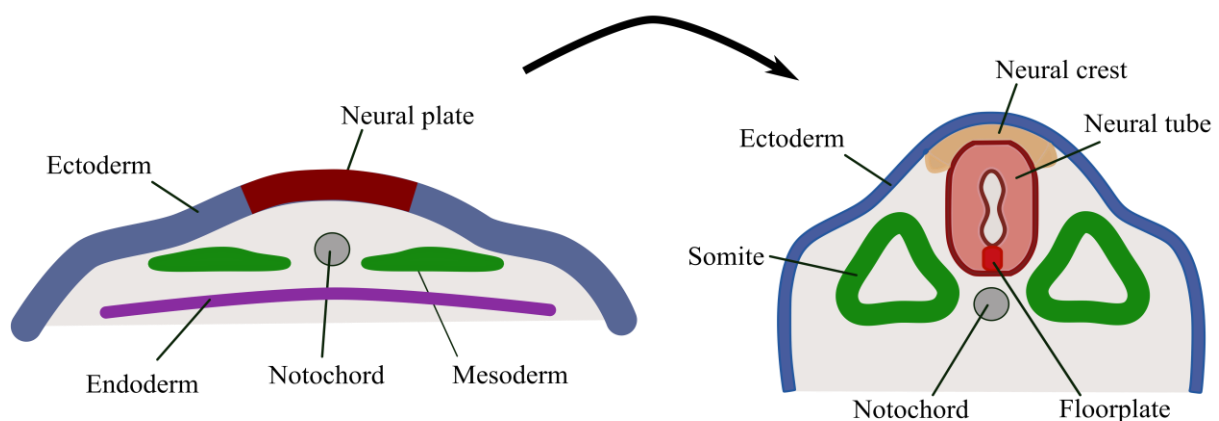
Figure 2: Schematic of neurulation. Adapted using Inkscape from Purves et al 2001.

Due to the expression of a combination of transcription factors surrounding the neural tube, the progenitor domain will form in the ventral spinal cord. Progenitor MNs (pMN) arise from the section of the domain which expresses a combination of genes including *NKX6.1*,

PAX6, and *OLIG2*. At this point the pMN is ready to exit the cell cycle and begin the differentiation process to become a mature MN[6]. The process is highly regulated in order to produce the correct amount of MNs at the appropriate time. Among other factors influencing this complex mechanism, OLIG2 contributes to the expression of neurogenin 2 (*NEUROG2*). This is an especially interesting interaction, where in a precise coordination of timing and concentration, OLIG2 appears to hold pMNs in a dividing state, but also subsequently opens the course for NEUROG2 to push the developing motor neuron into the differentiated state[4]. NEUROG2 interacts with the RA receptor and influences the transcription of MN genes, including one important downstream target, the motor neuron and pancreas homeobox 1 (*HB9*, aka *MNX1*). Once expressed, HB9 can self-regulate and is therefore independent from other pMN signals and, along with Islet1, is a reliable marker for post-mitotic MNs[6].

Despite the complexity of the system only a relatively small number of factors are actually needed to modulate cell fate and development, these developmental cues occur at specific times and concentrations and in certain combinations[7]. Therefore, the derivation of MNs as an *in vitro* model system should theoretically be possible. However, the differentiation of pluripotent cells into motor neurons still has many hurdles to overcome. Multiple groups have taken on the challenge to develop technical protocols to guide the efficient production of pure and functionally mature MNs resulting in reported populations ranging from 70 to 90 %[8,9]. These MN populations can be derived from patients with specific MN diseases, such as amyotrophic lateral sclerosis, in order to study patient and disease specific changes in the cells[10]. The populations can also be derived from iPSCs from a healthy donor, which can be genetically modified to investigate a specific hypothesis based upon the modified cells. Additional aspects about the relevance of iPSCs are presented in more detail below.

1.1.2 The Neuron

Neurons are essential for the organism to process and integrate responses to the outside world. While the diverse population of cells found in the human body can share many features, neurons have more complex cell specializations than any other cell type[11]. Each specific neuronal type has a characteristic structure and neurons can range in size from just a few to greater than 100 micrometers (µm)[12]. Neurons contain a large-scale translational cytoplasm consisting of polysomes, rough and smooth endoplasmic reticulum, and the Golgi complex, responsible for the high protein synthesis and transport necessary for the neuron to function and prosper[11,12]. As illustrated in Figure 3, the cell body (soma) of the neuron containing the nucleus

4

narrows into the axon which can end after a couple of µm or extend to more than one meter in length. Secretory vesicles, vesicles containing secretory polypeptides, as well as neurotransmitter elements are transported from the translational cytoplasm towards the axonal termination with the help of microtubules and motor molecules by fast anterograde axonal transport. These axonal terminations form the site of the synapse where the arriving action potential triggers the opening of voltage-gated calcium channels, resulting in the fusion of synaptic vesicles with the presynaptic terminal membrane and the release of the neurotransmitter into the synaptic cleft. The neurotransmitter binds to the postsynaptic receptors and the signal propagates[12].

The main neurotransmitter of lower motor neurons, found in the spinal cord and enervating outside the central nervous system, is acetylcholine (ACh)[6]. ACh is synthesized by choline acetyltransferase from acetyl coenzyme A and choline[12]. The vesicular ACh transporter (VAChT) actively transports acetylcholine into the SVs of cholinergic motor neurons[13]. Upon depolarization of the motor neuron, the SVs fuse with the plasma membrane, the ACh is released into the motor neuron end junction where it can bind to acetylcholine receptors along the postsynaptic muscle fiber membrane. This interaction triggers a series of depolarizations and finally a muscle fiber action potential and contraction. Remaining ACh is hydrolyzed by acetylcholine esterase and the cycle of synaptic transmission is complete[14].
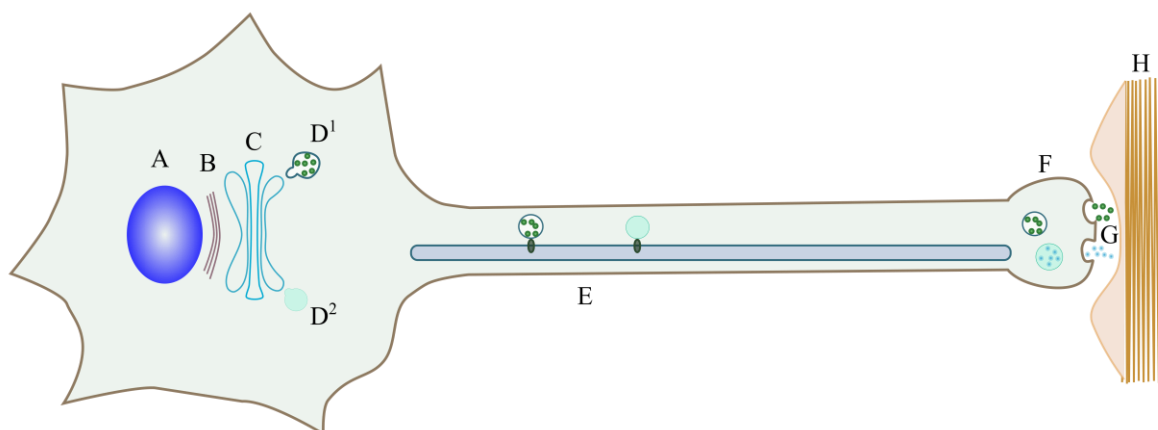
Figure 3: Simplified illustration of a motor neuron and selected components A) nucleus, B) Endoplasmic reticulum, C) Golgi apparatus, D^1) large dense core vesicle with cargo, D^2) synaptic vesicle E) large dense core vesicle and synaptic vesicle carried to presynaptic terminal by fast anterograde axonal transport, F) vesicles reach presynaptic terminal, acetylcholine is loaded into synaptic vesicle, G) synaptic cleft (general designation) or motor neuron end junction (motor neuron), H) muscle fiber; figure adapted using Inkscape from Purves et al 2001.

1.2　Neurosecretory Pathway

1.2.1　Neurosecretory vesicles

Neurosecretory vesicles, encompassing both large dense core vesicles (LDCV) and electron-lucent synaptic vesicles (SV), are present in all neurons and have evolved to deliver cargo, such as peptides and proteins, to the exterior of the cell. The secretory cargo proteins contain an N-terminal leader sequence that warrant co-translational translocation of the nascent protein into the rough endoplasmic reticulum. Additional sorting signal sequences target these proteins into specialized vesicles. The proteins are then transported through the Golgi stack ending in the *trans*-Golgi network (TGN) where the proteins are sorted into immature vesicles which bud off the TGN. The membrane of the vesicle begins to pump protons into the lumen, which reduces the intravesicular pH, concurrently maturing the neuropeptide molecules and stabilizing the vesicle within the cytoplasm[15]. The classical neurotransmitter is typically synthesized in the axon and can enter and accumulate in the SV with the help of specific transporter proteins. Peptide neurotransmitters, on the other hand, typically are derived from prohormones which passed through the Golgi apparatus and are packaged in budding LDCVs[16].

While the transport through the Golgi to become a SV is not as well characterized as that for the LDCV, some parallels and diversions in the trafficking of vesicle proteins through the Golgi have been observed. In an experiment assessing the colocalization of two proteins associated with the SV, VAChT and synaptophysin (Syp), and the LDCV associated protein chromogranin A (CgA), the two synaptic vesicle proteins constantly and significantly colocalized. On the other hand, the SV proteins only intermittently colocalized with CgA. For example, both Syp and CgA were found to move through Golgin97-positive compartments of the Golgi, whereas only Syp and VAChT, but not CgA, could be found in TGN46-positive compartments of the Golgi, indicating the proteins for each vesicle type can be processed in separate sub-compartments[17]. Once the vesicles bud from the TGN, the membrane and secretory proteins remain associated with the vesicle for the lifetime of the cell[18]. Classic proteins that are found on the SV membrane, and are therefore markers for SVs, are involved in neurotransmitter transport, membrane fusion and Ca^{2+} signaling, such as VAChT, vesicle-associated membrane protein (VAMP, aka synaptobrevin), and synaptotagmin, respectively[19]. LDCVs can contain multiple proteins, such as prohormones like proopiomelanocortin (POMC), or Ca^{2+} binders in the granin family such as CgA and secretogranin II (SgII)[20,21].

1.2.2 Membrane fusion

Classical neurotransmitters and neuropeptides can be released at separate timepoints, even within the same cell, implying different mechanisms might play a role in the secretion of each neurotransmitter type from their specific vesicle-type[16]. For example, it has been shown in *C. elegans* that protein kinase C-1 is necessary for release of NLP-21 neuropeptide secretion from LDCV in cholinergic motor neurons, while SV fusion can still occur in the absence of the protein[22]. Furthermore the release of SVs is very rapid after a single action potential, taking only a few hundred microseconds, whereas the LDCVs require successive stimuli resulting in a delay of release in the range of milliseconds[23]. Despite the difference in conditions required for fusion of either the SV or LDCV, both types of vesicle require the SNARE (soluble N-ethylmaleimide-sensitive factor (NSF) attachment receptor) complex for fusion to the presynaptic membrane and secretion of their contents into the synaptic cleft[24].

SNARE protein complexes are made up of various isoforms of VAMP 1 and 2, syntaxin, and synaptosomal-associated membrane protein (SNAP-25). These proteins selectively interact with each another to form SNARE complexes, which are necessary for the fusion of the vesicle to the presynaptic membrane[24]. Also crucial for intracellular membrane fusion is the interaction of the SNARE complex with Sec1/Munc18-like proteins[25,26]. Initiation of the fusion process begins with the depolarization of the cell membrane, and the subsequent Ca^{2+} influx and binding to the calcium sensor synaptotagmin-1 (Syt1) (step 1 in Figure 4). The Munc18 protein then binds to syntaxin-1, whose closed conformation normally blocks access to the protein's SNARE motif. Munc18 binding changes the conformation of syntaxin-1 to an open form, in which the protein is actively available to other members of the SNARE protein family (step 2 in Figure 4). VAMP, a vesicular transmembrane protein, can then interact with syntaxin-1 and the transmembrane protein SNAP-25, both of which are located on the plasma membrane. Munc18 remains associated with syntaxin-1, which is essential to the stability and function of the SNARE complex (step 3 in Figure 4). The helical structures formed by the complex progressively zipper together, forcing the membranes of the vesicle and synaptic terminal closer and closer, eventually disrupting their hydrophilic surfaces and opening the fusion pore[27] (step 4 in Figure 4). The expansion of the fusion pore changes the conformation of the SNARE complex, which has lost its potential energy and can only dissociate with the help of NSF (an ATPase)[25,27]. Endocytosis of the synaptic vesicle occurs and the cycle of fusion is complete.
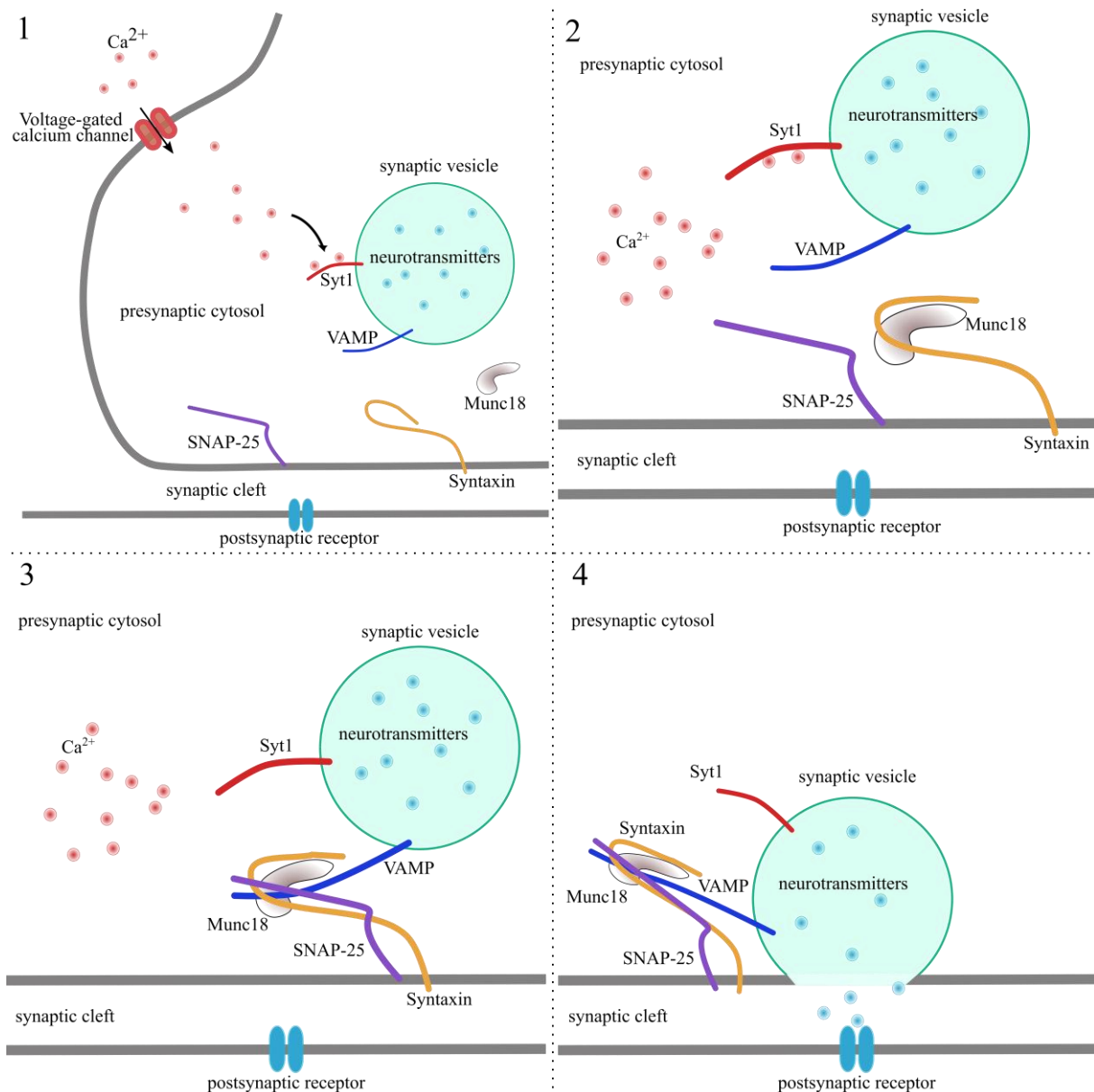
Figure 4: Simplified schematic of SNARE complex mediated vesicle fusion at presynaptic membrane. (1) A stimulus causes the influx of Ca^{2+} into the presynaptic terminal, Ca^{2+} binds to the calcium sensor synaptotagmin-1. (2) Munc-18 binds to syntaxin-1, changing it into an open conformation. (3) VAMP and SNAP-25 can interact with the open conformation of syntaxin-1, forming the SNARE complex. (4) SNARE complex zippers together, forcing the fusion of the vesicular membrane and the plasma membrane. Image created with Inkscape.

1.3 Botulinum Neurotoxins

1.3.1 Background

Botulinum Neurotoxins (BoNTs) are proteins produced by anaerobic bacteria of the genus *Clostridium* (e.g. *Clostridium botulinum*) and are among the most noxious substances known in the world. The lethal human dose by injection is approximately 1 ng/kg and by ingestion is 1 µg/kg[28]. The harmful effects of BoNT are caused by the toxin's ability to enter neurons where it cleaves members of the SNARE complex, resulting in the impairment of

8

neurotransmitter release and signal propagation. While this action is entirely reversible, because of the stability of certain forms of BoNT, the poisoning of the nerve cell and blockade of neurotransmitter release can last for multiple months[29]. Accidental exposure to the toxin through various routes including ingestion, anaerobic wounds or inhalation can result in botulism, causing paralysis by inhibiting acetylcholine release at peripheral nerve terminals. All forms of botulism are characterized by descending paralysis starting with the ocular muscles, then expanding through the facial muscles to the respiratory muscles, concluding in respiratory failure[30].

In the 1970s, a collaboration between Alan B. Scott, a surgeon at the Kettlewell Eye Research Institute in San Francisco, and the research group of E. Schantz resulted in the successful treatment with BoNT of induced strabismus in rhesus monkeys, leading to the treatment of the visual disorder in human volunteers in the 1980s. The use of BoNT was considered to be optimal, given its effectiveness over many weeks, however it became obvious that great care must be used to prepare the toxin for commercial use in humans[31]. Close attention must be paid to the purification process and the final dosage issued to patients. The application of BoNT in medical and cosmetic fields has expanded considerably since the 1980s, treating ailments in neurology as well as urology, pain reduction, and hyperhidrosis[32]. BoNTs used as clinical formulations are manufactured by pharmaceutical companies. To this end, the bacteria is fermented and the neurotoxin protein is isolated, for example by acid precipitation followed by column chromatography[33]. While these techniques are highly regulated, it is necessary to verify the activity, stability, and quality of the extracted neurotoxin to continue to safely utilize the drugs in standard medical treatment.

1.3.2 Protein structure

BoNT is synthesized as a 150 kDa inactive polypeptide chain composed of three equisized 50 kDa domains. The toxin is activated by post-translational proteolysis, cleaving the protein by clostridial or tissue proteases into light (LC – 50 kDa) and heavy chains (HC – 100 kDa), which are linked by a disulfide bond[34–36] (Figure 5).

There are three main stages in the process of BoNT intoxication, each of which is controlled by one of the domains on the light and heavy chains[36]. The HC has two functional domains controlling internalization and membrane translocation. The C-terminal domain (H_C) determines the protein's affinity and specificity for neurons by binding to ganglioside receptors and surface membrane receptors, and therefore controls vesicular internalization of the toxin at the neuromuscular junction of motor neurons. Once internalized in the presynaptic terminal, the acidic milieu of the vesicle causes the N-terminal domain (H_N) to form a transmembrane channel which is responsible for translocation of the endopeptidase encoded in the functional domain of the LC from the vesicle into the cell cytosol. Upon reaching the reductant and neutral pH of the cytosol, the disulfide bond is reduced and the endopeptidase is released[34,35]. Standing alone, the LC would act as an ordinary protease. The HC is essential to the extreme toxic effects of BoNT, since it delivers the protease directly into the motor neurons to cleave the members of the SNARE protein complex[37]. Despite a similarity in structure, BoNTs are diverse proteins, consisting of at least seven different serotypes (BoNT/A to BoNT/G) and more than 40 subtypes[38]. The serotypes range in sequence similarity between 37 – 69 %[37]. Despite their differences, the common factor for each of these serotypes is their ability to target a specific site in the SNARE protein family[38].

1.3.3 Molecular mechanism

1.3.3.1 Binding and endocytosis

The first stage of BoNT intoxication is the membrane binding and internalization of the entire protein structure. Upon intake of the toxin in the body, BoNT enters the lymphatic and blood circulation, arriving at the perineuronal fluid compartment. BoNT binds with high affinity to the presynaptic plasma membrane of skeletal and autonomic cholinergic nerve terminals[30]. The process of internalization is quick, with most of the toxin likely being taken up into the SV itself during vesicular endocytosis[39]. The dual receptor hypothesis postulates that the H_C binds to a presynaptic polysialo-ganglioside (PSG) receptor and then to a protein receptor (Figure 6). The gangliosides bind at the domain's highly conserved ganglioside-binding site[40].

Detailed information is still lacking for some of the BoNT serotypes, but the known secondary protein receptors at the time are synaptotagmin (Syt), which binds the serotypes

10

BoNT/B, DC, and G, and the glycosylated synaptic vesicle protein 2 (SV2), which binds the serotypes BoNT/A, D, E, and possibly BoNT/F[30,40]. The hypothesis that the internalization of the BoNT/A appears to be modulated by the SV2 is supported by the detection of 1-2 molecules of BoNT/A per synaptic vesicle, which matches the estimated copy number of SV2 molecules per vesicle[19]. The stimulation of the nerve increases the uptake and thus toxicity of BoNT, which also indicates the toxin for treatment of hyperactive nerve terminals, increasing uptake and effectiveness at these specific targets[30].
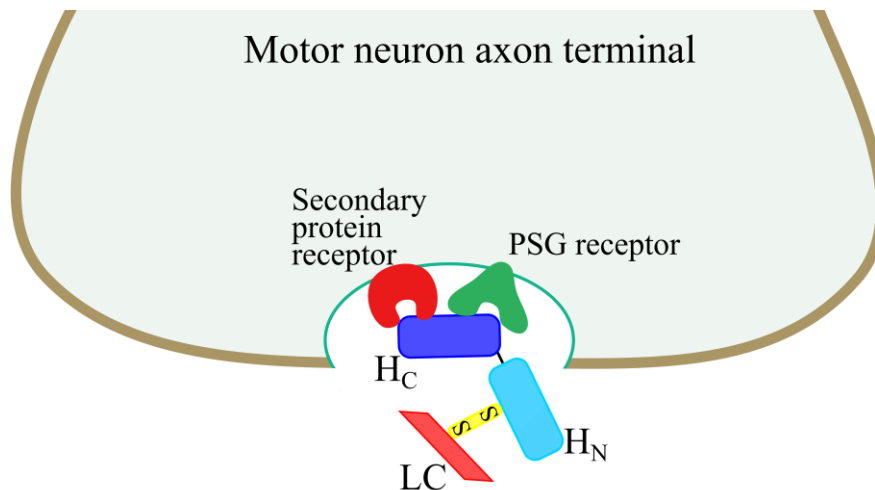
Figure 6: Illustration of BoNT internalization into motor neuron axon terminal by interaction of the C-terminal heavy chain (H_C) with the polysialo-ganglioside (PSG) receptor and a secondary protein receptor during vesicular endocytosis. Cellular elements are not drawn to size.

1.3.3.2 Exocytosis/Membrane translocation

Upon re-internalization of the SV from the presynaptic membrane, the vesicle automatically begins the recycling process to re-internalize neurotransmitters and prepare for renewed exocytosis at the end junction. The most important step in this process for the BoNT protein is the reacidification of the vesicle (step 1 in Figure 7). Translocation of the LC from the vesicle to the cytosol takes place once the lumen of the SV reaches between pH 4.5 - 6, indicating that the SV must be mature for the toxin to reach its full proteolytic capacity[30]. The reduced pH is thought to provoke a structural change in BoNT, increasing its hydrophobicity[36]. Once the environment in the SV acidifies, the H_N domain can act as a chaperone for the LC to move through the SV membrane, into the cytosol, at which point the toxin is rapidly reduced (steps 2 and 3 in Figure 7). The two most likely possibilities for the translocation are the formation of a conduction channel by the H_N or the toxin may form a "molten globule", which would interact with the hydrophobic luminal membrane and anionic lipids, and deliver the toxin to the cell cytosol[30]. In either of these two options, the LC quickly translocates from the SV lumen into the cell cytosol and its endopeptidase properties are activated[39].
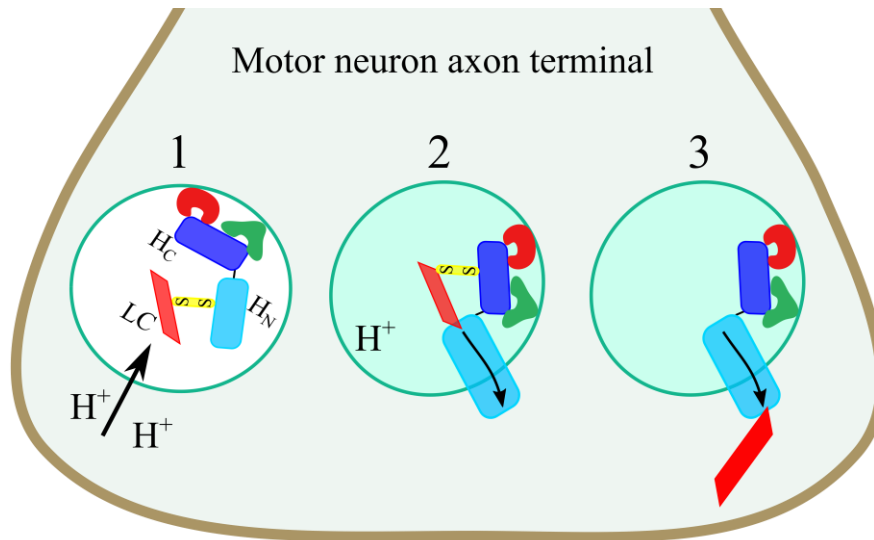
Figure 7: Illustration of steps necessary for translocation of BoNT light chain from vesicular lumen into cytosol. 1) Endocytosis of vesicle including BoNT from the presynaptic membrane, reacidification of vesicular lumen begins. 2) Vesicular lumen is acidic (indicated by green colored lumen) and the N-terminal heavy chain domain forms a pore in the vesicle membrane. 3) The light chain translocates from the vesicular lumen into the cytosol of the motor neuron axon terminal. The disulfide bond connecting the light chain to the heavy chain is reduced and the light chain becomes an activated protease.

1.3.3.3 Cleavage and consequential neuroparalysis

In the process of normal neuronal signal propagation, the fusion event of the SV to the presynaptic membrane results in the release of neurotransmitters into the synaptic cleft and the propagation of the signal to the next neuron or muscle fiber. Upon activation and entry into the cytosol of the motor neuron presynaptic terminal, the enzymatic domain of BoNT can cleave and inactivate specific cellular proteins in the SNARE family, blocking neurotransmitter release. In the presence of each of the BoNT serotypes, a specific peptide sequence is recognized and cleaved within the respective SNARE substrate. Each individual cleavage event can prevent the formation of a stable neuroexocytosis apparatus. The resulting inhibition of signal propagation results in botulism, a flaccid paralysis of the skeletal and autonomic nervous systems[30].

Each BoNT serotype targets a specific substrate, each a member of the SNARE family. VAMP is an integral membrane protein of SVs and LDCVs and is subject to cleavage by BoNT/B, D, F, and G. Syntaxin 1 and 2 are present in the presynaptic membrane of neurons and subject to cleavage by BoNT/C. SNAP-25 is a highly conserved protein on the cytosolic side of the neuronal membrane, which appears to interact with syntaxin to form a complex that may act as a VAMP receptor. SNAP-25 is subject to cleavage by BoNT/A, C and E. Under normal physiological circumstances a single isoform of each of the three proteins associate, bringing the vesicle and presynaptic plasma membrane in close proximity, enabling membrane

12

fusion. Upon exposure to BoNT the complex cannot be formed and neurotransmitter release is blocked[24]. The mechanism of BoNT entry into the motor neuron and cleavage of SNARE proteins is summarized in Figure 8.
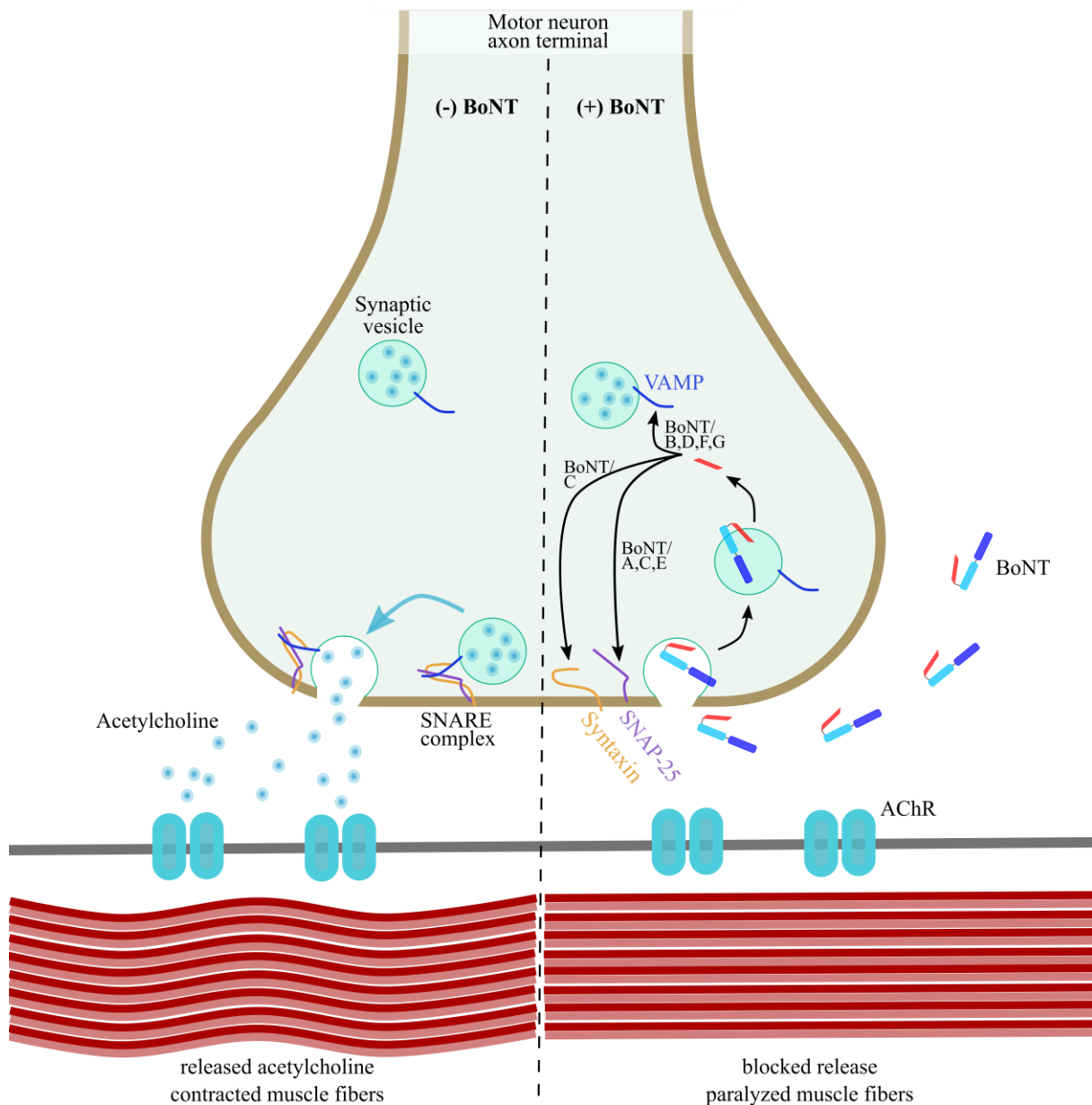
Figure 8: Pathway summary of BoNT in the motor neuron. Normally functioning MN on left side (- BoNT), MN exposed to BoNT on right side (+ BoNT). Figure adapted using Inkscape from Singh et al 2013.

1.3.4   Toxicity assessment

1.3.4.1   Replacement, Reduction, Refinement

The use of animals in the laboratory is varied and controversial. The purposes range from drug or toxicity testing to the general advancement of knowledge[41]. Important and useful information can be obtained with the help of animal testing. However, the ethical issues and

actual pertinence and transferability to humans have been questioned and discussed in the scientific community for over 60 years. In 1959 Russell and Burch published their book *The Principles of Humane Experimental Technique*, introducing the concept of the three Rs (Replacement, Reduction, and Refinement) for the first time[42]. Specifically, the three Rs suggest the substitution of conscious animals for insentient material; the reduction in the number of animals necessary for the information needed through improved and more concise experimental design; and for those animals which are unavoidable to use, an improvement in the inhumane procedures applied to them[41].

1.3.4.2 The mouse bioassay

The "gold standard" in BoNT potency testing

The traditional method for potency determination of BoNTs is the mouse bioassay (MBA, also known as the mouse lethality assay) measuring LD_{50} potency, also known as the mouse lethality test, currently considered the "gold standard" in BoNT toxicity testing. A valid MBA must contain a range of doses spanning from a 90 % death rate to a 90 % survival rate, the precision of the test is dependent on the number of dilutions used and the number of animals exposed to each dosage. The dosing generally causes respiratory failure because of the paralysis of the respiratory muscles, and therefore extreme suffering on the part of the animals[43]. At higher doses, mice can show symptoms of botulism within 8 hours, however at lower doses, onset is much slower and animals must be observed for 4 days before establishing a negative result[44].

Despite the fact that this assay is well established and accepted by regulatory bodies, it has its own problematic beyond the ethical concerns of animal testing. The assay can have a large margin of error, involves large numbers of mice, and necessitates specialized animal facilities and staff[45]. Furthermore, the reaction of mice to unique BoNT serotypes does not always correspond 100 % to the human response, for example BoNT/B demonstrates a much lower potency in humans than mice. It has been shown that the human SV protein receptor, synaptotagmin-II (Syt-II) contains an amino acid change which may impair BoNT/B's recognition of its high-affinity binding site in comparison to that of the mouse[46]. The induced suffering of the creatures, the complications of working with them, and the natural differences between animals decreases the applicability of the MBA and pinpoints the need for novel alternatives to toxicity testing of BoNT in mice.

14

1.3.4.3 The future of BoNT potency testing

Due to the influence of the three R goals, both government and research entities are motivated to develop alternative methods to accurately and precisely establish the potency of BoNT lots. Furthermore, the problematic of inter-species variation increases the relevance of establishing human specific cell-based assays. The development of alternative assays to the MBA, competitive in their sensitivity to specific forms of BoNT, began as early as 1999, with the *in vitro* assay for the detection of BoNT/B with the help of a modified antibody enzyme-linked immunoassay (ELISA) detecting cleavage products after exposure to the toxin[47]. Since then, a variety of tests have been developed, each with their own pros and cons. The newly developed detection methods can be divided into following categories[48]: the *in vivo* MBA, which can be improved by local injection methods, but which still uses animals and requires specially trained personnel[45]; *in vitro* (cell-free) assays, such as the immunological detection method ELISA, which can rapidly detect the presence of toxin but are generally unable to distinguish between active and inactive forms, and endopeptidase assays measuring the enzymatic activity of the LC, but which cannot distinguish between fully active holotoxin and LC; *in vitro* (*in vivo* simulation) assays such as cell-based assays using primary or immortalized cell lines. Furthermore, *ex vivo* assays have been developed, which measure the contraction of dissected indirectly stimulated muscle. See Table 1 for summary of assays used for BoNT potency testing.

Table 1: Summary of selected methods to assess BoNT toxicity.

Category	Test method	BoNT serotype	Detection method
in vivo	Mouse bioassay	all	90% asphyxiation / survival rate
ex vivo	Mouse Phrenic Nerve Hemidiaphragm Assay	BoNT/A,B,E[49]	measuring contraction amplitude of indirectly stimulated muscle
***in vitro* (*in vivo* simulation)**	hiPSC-derived neurons	BoNT/A[48,50,51]	SNAP-25 ELISA / WB
	Primary rat spinal cord assay	BoNT/A,B,E[52]	Western blot
	PC12 cells	BoNT/A,E[53]	FRET - GFP/RFP transfected
	NG108-15 Neuronal / S16 Schwann cell coculture	BoNT/A[54]	SNAP-25 cleavage in western blot
	SIMA neuroblastoma cell line	all BoNTs[2]	co-release of transfected luciferase from LDCVs upon stimulation

	SIMA neuroblastoma cell line	BoNT/B[55]	Nano-Glo ELISA treated VAMP-GFP modified cell sample added to attached cleaved antibodies
	mouse embryonic stem cell-derived neurons cultured on multi-electrode arrays	BoNT/A[56]	spontaneous network bursts
in vitro **(cell-free)**	ELISA	BoNT/B[47], BoNT/A[57], BoNT/A,B,E,F[58]	antibody recognizes cleaved target
	Mass spectrometry	BoNT/A,B,E,F[44]	endopeptidase activity by detecting cleavage products
	Endopeptidase assay	BoNT/A,B[59]	antibody recognizes cleaved target

These *in vitro* (cell-free) methods are a large step forward in the assessment of the full biological activity of BoNT, however, they are all hindered by the inflexible characteristic that they only detect one specific serotype of BoNT, since they measure the accumulation of the cleavage product of that specific BoNT serotype. The *in vitro* (*in vivo* simulation) assays are those with the greatest potential to provide the most reliable and complete potency estimations. These are capable of incorporating all steps of the intoxication process: the uptake of the full-length protein in the motor neuron endplate, the vesicle formation, the transduction of the LC from the vesicle into the cell cytosol, and finally, cleavage of the appropriate SNARE protein. Only if all these factors are considered is it truly possible to make an accurate statement about the full activity of the BoNT.

There are several assays which have been developed using neuronal or immortalized cell lines, one example, the SIMA Random_Insertion-hPOMC1-26GLuc prototype[2] upon which this project is based, measures the actual endpoint of the functional neuron: the release of neurosecretory vesicles from the presynaptic terminal. This secretion normally takes place under stimulated conditions, however, if any members of the SNARE proteins are exposed to and cleaved by any BoNT serotype, the neurosecretory vesicle release will be inhibited. Therefore this assay has the great advantage that it should be able to flexibly measure the outcome of exposure to any of the seven main BoNT serotypes. This overall inhibition is measured by the decrease of Gaussia luciferase co-released with neurotransmitters into the supernatant surrounding the SIMA cells[2]. The drawbacks of this assay include the fact that the cells are not the perfect natural target of BoNT, since they are an immortalized neuroblastoma

cell line as opposed to cholinergic motor neurons. Furthermore, they have been transfected with a method that inserts the reporter gene into the genome in a non-targeted and uncontrollable manner, which could result in either an insertion of the donor DNA in a gene that is important for normal cell function or in multiple insertions of the donor DNA throughout the genome.

1.4   Induced Pluripotent Stem Cells

The first multipotent stem cells derived from bone marrow cells were identified in 1961[60] followed by the discovery of pluripotent cells in early mouse embryos 20 years later[61]. Researchers determined that these cells retained their ability to develop into a variety of germ layers, and could furthermore maintain the stem cells *in vitro* on feeder cell lines. Taking advantage of site-directed mutagenesis in the embryotic stems cells, it was possible to generate mouse models with a particular phenotype, for example to model human disease. The first human embryonic stem cells were generated from blastocytes and could be maintained in an undifferentiated state for months in cell culture[62,63]. The concept of pluripotency was at the forefront of the biological field and soon researchers began experimenting with various transcription factors which might potentially play a role in stem cell pluripotency.

In 2006 the group of Shina Yamanaka narrowed down four transcription factors which were found to be necessary and sufficient to revert mouse embryonic fibroblasts to a pluripotent state[64]. This discovery opened the doors to the development and use of induced pluripotent stem cells (iPSC), making stem cell research ethically viable and more technically accessible. It was no longer necessary to obtain stem cells from an embryo, which had been of ethical concern. The method to induce the pluripotent state was also relatively simple, involving protocols using small molecules, microRNAs, and combinations of reprogramming factors. The technology opened the doors to potential treatment using patient-specific cells as well as a powerful method to model disease, which could be more accurately and easily achieved with the right genetic editing tools[65].

1.5   A short history of genetic manipulation

Genetically modified cell lines and animals have been used for decades to advance scientific knowledge by making specific modifications to the genome of the model of interest. Already in the 1960s it was shown that viral DNA could be stably integrated into target cells. Using a calcium phosphate transfection method in the 1970s, researchers could modify approximately 1 in 100,000 cells[66]. Since then genetic engineering technologies have improved

immensely both in efficiency and precision. The idea that the integration events occurring during the transfections could be due to homology-directed recombination (HDR) was introduced in the early 1980s, suggesting that targeted insertions of donor DNA into a specific genomic locus might be possible[67]. In 1985 the insertion of a DNA insert into the human β-globin locus was achieved by electroporation in 1 out of 1000 cells[68]. Genetic editing was becoming more precise, with a higher frequency of successful events.

In order to specifically target a genomic locus for manipulation it is necessary to instigate a double-stranded break (DSB) at the point of interest with a nuclease. The natural DNA repair mechanisms of non-homologous end-joining (NHEJ) and HDR can then be exploited to either insert a donor DNA sequence, cause a mutation, or remove a segment of genomic DNA. The progress towards development of site-directed mutagenesis tools also got off to an inefficient start. However, specialized systems such as meganucleases, zinc finger nucleases (ZFNs), and transcription activator-like effector nucleases (TALENs) have been developed, which have improved the possibilities of genetic manipulation. These systems vary, however, in cost and ease of production[62].

1.5.1 CRISPR/Cas9 technology

In 2012 the CRISPR/Cas (cluster, regularly interspaced, short palindromic repeats/CRISPR-associated) method was introduced to great acclaim, winning the discoverers of the system, Emmanuelle Charpentier and Jennifer A. Doudna, the Nobel Prize in Chemistry in 2020. The technology takes advantage of the natural immune mechanism of many bacteria against viruses and other foreign nucleic acids by which the foreign DNA is specifically detected and silenced by small RNAs produced within the host. As the name suggests, this system is made up of two parts. The CRISPR domain contains a sequence specific to the invading DNA (protospacers) inserted into a cluster of short repetitive sequences in the host's genetic makeup. The *cas* genes encode endonucleases, which can be directed by elements of the CRISPR domain to cut at a specific protospacer target, the foreign DNA in the case of the immune response[69]. In its application as a tool in genetic editing, an active endonuclease (Cas9) can be directed to a specific site in the genome to induce a double strand break (DSB) at that point. The active components of the system are[70] (illustrated in Figure 9):

A. The Cas9 enzyme: the endonuclease responsible for cutting the gDNA.

B. The crRNA array: encoding the guide RNA and the transactivating crRNA (tracrRNA). These are commonly fused together to create a single-guide RNA (sgRNA), the 20 nucleotide guide sequence directs the Cas9 to the gDNA target.

C. The protospacer: in the microbial immune system this is the exogenous DNA target, in the context of genetic editing this is a 20-nt sequence of the target gDNA where the DSB will be induced.

D. The protospacer adjacent motif (PAM): a sequence associated with the protospacer, which is specific to unique Cas9 orthologs and assists in the direction of the sgRNA to the protospacer sequence.
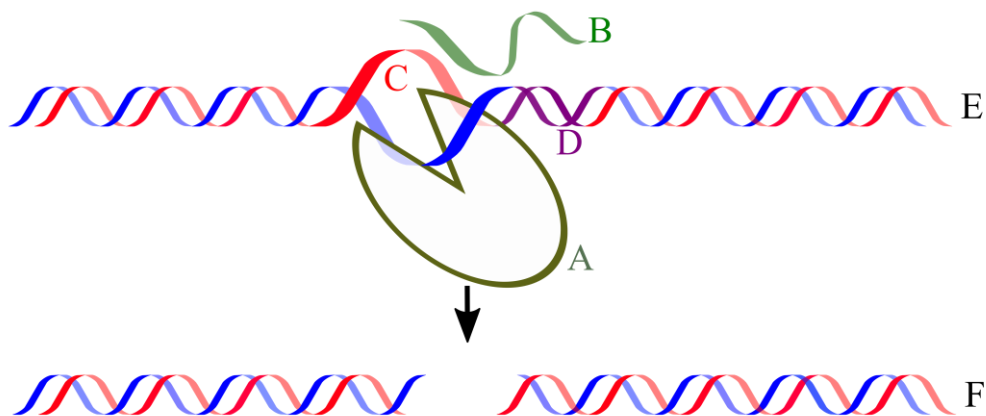
Figure 9: Components of the CRISPR/Cas9 genomic editing system. (A) Cas9 Endonuclease, (B) crRNA array aka sgRNA, (C) protospacer, (D) PAM, (E) target DNA, (F) target DNA with double strand break.

Cas9 is directed to a specific site in the target gDNA by an sgRNA, which is recognized by the matching Watson-Crick base pairing of the sgRNA to the protospacer sequence of the gDNA. The Cas9 endonuclease induces a DSB at this site in the target gDNA. After a DSB occurs, the cell's natural DNA repair mechanisms kick in, which can be taken advantage of in order to genetically modify the cell in a controlled way. The two major DNA repair mechanisms are homology-directed recombination and non-homologous end joining (Figure 10). The high fidelity mechanism of HDR can be utilized to cleanly insert modifications into gDNA with the help of a repair template consisting of the donor DNA surrounded by homology arms (HA) pertaining to either side of the DSB. NHEJ is an error-prone mechanism in which each end of the DSB are ligated back together, often producing mutations which could lead to a dysfunctional protein or a premature stop codon[70].
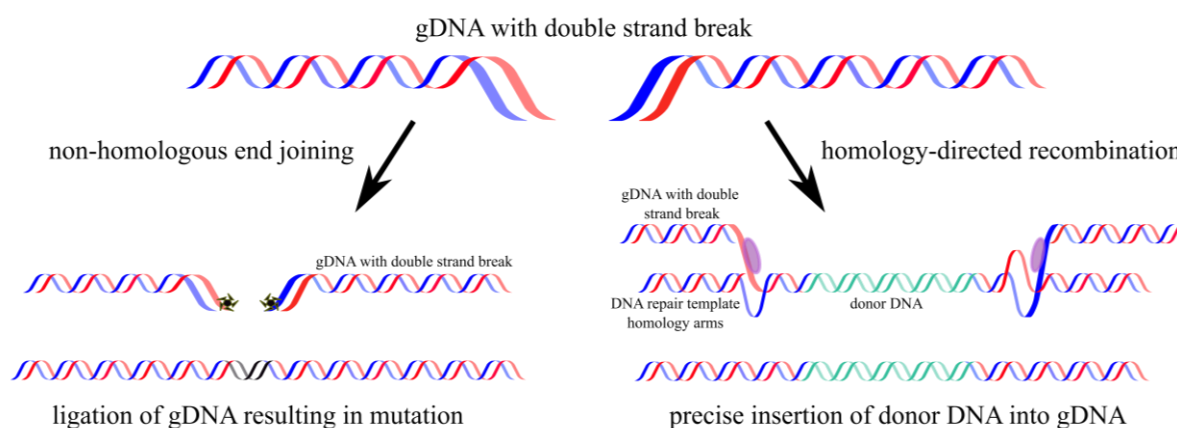
Figure 10: Graphic representation of DNA repair mechanisms HDR and NHEJ. Adapted using Inkscape from Ran et al 2013.

1.5.2 Genomic safe harbors

CRISPR/Cas enables researchers to manipulate gDNA at specific loci, allowing the manipulation of specific genes of interest. The question remains, however, how to incorporate external donor DNA into an organism's gDNA without disrupting other genes necessary for normal cellular function. To solve this problem, loci called "genomic safe harbors" (GSH) have been proposed. GSHs can possibly be located in non-essential genes, intragenic sites within gene-rich areas, or extragenic sites. No GSH has been completely validated, but there are currently three candidate safe harbors which have already been tested for targeted donor insertions[71]:

1. AAVS1: The adeno-associated virus site 1 is located on chromosome 19 in the PPP1R12C gene. It is site of repeated AAV integration in infected human cell lines *in vitro*. High expression of transgenes could be observed in iPSCs in all three embryonic germ layers post-differentiation and no abnormalities were observed in those cells with genetic manipulations at this site.

2. CCR5: The chemokine (CC motif) receptor 5 is located on chromosome 3 and is a major co-receptor for HIV-1. Disruption at this site in the form of a naturally occurring mutation confers HIV resistance, but is not associated with any other major pathology. Reporter genes integrated into this site display lower expression levels than those integrated in the AAVS1 site.

3. Human ROSA26: Human orthologue of the mouse ROSA26 locus is located on chromosome 3. Expression of reporter genes integrated at this site remains after differentiation in all three germ layers, however few studies have confirmed the safety or utility of this locus.

1.5.3 CRISPR/Cas and iPSCs – a revolution in genetic research

An initial challenge faced in the use of human pluripotent stem cells (hPSCs) in research was the unexpected difficulty in genetically modifying the cells, especially in comparison to previously developed modification protocols using mouse embryonic stem cells which were easier to implement. Protocols to modify hPSCs were developed, but were very time consuming and inefficient, hindered even further by low single-cell survival rates. However, many advances have been made in the field. The incorporation of Rho-kinase inhibitor when splitting the cells in culture greatly increases the survival rate of manipulated cells. Site-specific nucleases have transformed the ability to use homology-directed recombination to efficiently modify genomic regions[65]. The potential application of accurate and precise targeted genome editing in pluripotent human cells has opened many doors analyze and possibly treat genetic disorders.

1.6 Goals and Strategies

The iPSC line IMR90-4 is used as the target of the CRISPR-mediated genetic engineering to prepare the MoN-Light BoNT assay. The aim of the modifications is to integrate a selection of constructs containing the genetic sequence for the Gaussia luciferase reporter protein associated with targeting sequences for LDCVs and SVs into the AAVS1 safe harbor. To attain this goal, a series of plasmids must be prepared for a CRISPR/Cas9 double-transfection, providing the Cas9 endonuclease and the sgRNA guide for the double-strand break, and the homology arms and donor DNA template for the homology-directed recombination. Post-transfection, cells in which the HDR has taken place and which have incorporated the donor DNA are selected for their consequential antibiotic resistance. The resulting clones must be fully characterized before subjecting them to the time and resource intensive differentiation into motor neurons and implementation of the luciferase release assay.

As previously alluded, the most important aspects of the clones to verify are the successful execution of the CRISPR/Cas9 genetic modification through analysis of donor DNA integration in the gDNA, and confirmation of the correct sorting pattern of GLuc into neurosecretory vesicles. A selection of techniques that can be applicable for the clone validation are explained in detail in the following sections. The methods that can be used to validate the modification of the cells fall into two categories. First, providing evidence that the donor DNA is integrated into the AAVS1 safe harbor locus. Second, verifying that the donor DNA is not integrated into any off-target genomic loci. The correct insertion of the donor DNA can be

analyzed with a basic PCR, amplifying the region surrounding the AAVS1 safe harbor locus and visualizing the resulting products on an agarose gel. There are several possible methods to evaluate the possibility of off-target donor DNA integrations. First, the long-standing Southern blot can be utilized to identify segments of digested DNA that include a specific sequence of interest. Second, the ligation-mediated PCR can be designed to amplify unknown regions of gDNA containing the donor DNA by using adapters ligated to digested gDNA ends and a known part of the insert sequence. Finally the copy number of a particular sequence in gDNA can be analyzed with qPCR. The localization of GLuc in the clones can be analyzed with differential centrifugation, separating cellular elements with increasing centrifugation velocities, or immunofluorescence, tagging specific proteins in cells with fluorescent markers. Only after all of these clonal attributes have been characterized is the evaluation of the clones in the MoN-Light BoNT assay justified.

1.6.1 Creation of plasmids for CRISPR mediated genetic modifications

The eSpCas9(1.1) plasmid[72] contributes the sequences necessary to both encode the enhanced Cas9 protein, the nuclease which initiates the double strand break at a specific point in the gDNA, and to encode the sgRNA, which directs the Cas9 to that specific point. The aim of this project is to incorporate a donor DNA sequence into the AAVS1 safe harbor locus, therefore the sgRNA sequence for the well-described T2 site at the AAVS1 safe harbor locus can be used[73]. The second plasmid necessary in the co-transfection using the CRISPR/Cas9 genetic editing method is the pAAVS1-P-MCS donor plasmid[74]. This plasmid contains two homology arms, pertaining to the 5' ("left") and 3' ("right") sides of the AAVS1-T2 safe harbor locus. After the Cas9 nuclease makes the double stranded break, the innate DNA repair tool, homology-directed recombination, can match the homology arms found in the plasmid to the gDNA and can use the plasmid as a repair template resulting in the transfer of the DNA sequence between the homology arms into the gDNA (see Figure 9). The plasmid's designation as a "donor" originates from these events. In the pAAVS1-P-MCS plasmid, between the left and right homology arms, there is a sequence encoding puromycin antibiotic resistance, as well as a multiple cloning site into which sequence(s) of interest can be cloned. The sequences of interest in this project are the Ef1-HTLV promoter expressing a fusion of the sequence encoding the Gaussia luciferase enzyme and a selection of specific signal peptides which should guide the transport of GLuc into neurosecretory vesicles. The preparation of the plasmids to incorporate the appropriate sequences specific to this project can be completed with standard

22

PCR of established and novel GLuc constructs and cloning into the vectors described above. Due to the potential incorporation of mutations into the sequence during PCR and the importance of directionality of the promoter and all following sequences, each newly incorporated sequence in the vector should be verified by Sanger sequencing.

The origin of this project is to use the SIMA Random_Insertion-hPOMC1-26GLuc prototype assay as the groundwork for the MoN-Light BoNT assay development, specifically using the hPOMC signal tag fused with GLuc as the construct to deliver GLuc to large dense core vesicles. However, in order to broaden the probability of identifying a clone which both successfully passes the characterization criteria and can release luciferase from differentiated motor neurons, additional constructs should be prepared with various signal peptides/sorting proteins in order to have a more diverse selection of clones. Some of the plasmid constructs can be reutilized from the previous optimization of the SIMA Random_Insertion-hPOMC1-26GLuc prototype, using signal peptides directing towards the LDCVs (designated hPOMC-GLuc, CgA-GLuc, and SgII-GLuc). Furthermore a construct can be designed using VAMP2 to direct GLuc to SVs (VAMP2-GLuc) as well as one with no sorting tag associated with GLuc (no tag Gluc) as a negative control to more precisely characterize the clones. A summary of the expected GLuc localization upon transfection with the previously described plasmid constructs is found in Figure 11. The constructs made with sorting tags for hPOMC, CgA, and SgII should express GLuc in the lumen of large dense core vesicles (left panel). The constructs made of a fusion of VAMP2 and GLuc, attached together by a TEV protease recognition sequence, should express GLuc in the lumen of the synaptic vesicle, fused to the vesicular membrane (middle panel). The constructs made with the no tag GLuc should express GLuc overall in the cell (right panel).

Expected GLuc localization according to respective sorting tag:
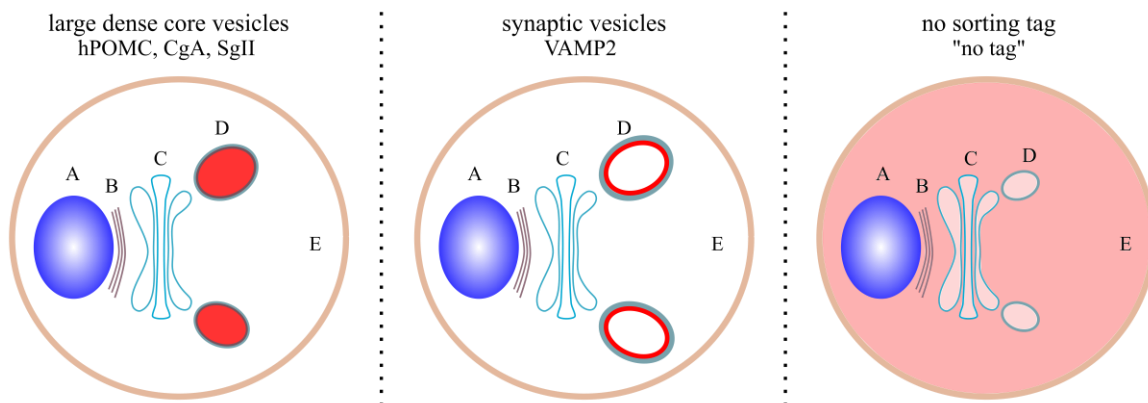
Figure 11: Summary of expected GLuc localization (depicted in varying shades of red) for each construct. A: nucleus, B: endoplasmic reticulum, C: Golgi apparatus, D: sprouting vesicle, E: cytosol.

1.6.1.1 no tag GLuc negative control

The GLuc construct with no sorting tag is included in this project as a characterization control and as verification of the specificity of luciferase release. The lack of a signal peptide means that the GLuc should be found in the cytosol of the cells, whereas all other tagged GLuc clones should sort the reporter enzyme through the ER and Golgi apparatus into LDCVs or SVs.

1.6.1.2 Summary of large dense core vesicle constructs

The hPOMC-GLuc signal peptide is the sequence encoding for amino acids 1-26 of the human proopiomelanocortin protein. The sorting signal designated as CgA contains a section of the 5'UTR and beginning of the *CHGA* genetic sequence. The SgII sorting sequence is designed analogous to CgA, consisting of the 5'UTR and beginning of the *SCG2* genetic sequence. The sorting signals directing towards the LDCVs were previously elucidated[2]. The luciferase release assay can theoretically be carried out with all of these constructs with the same protocol described for the Random_Insertion-hPOMC1-26GLuc prototype whereby cells are exposed to control and depolarization buffers and upon cellular depolarization the LDCV fuses to the presynaptic plasma membrane and releases neuropeptides and GLuc in parallel (Figure 12). The supernatant can be collected and GLuc can be measured by its reaction with the substrate coelenterazine, which undergoes oxidative decarboxylation, producing coelenteramide, CO_2 and light[75].

The prototype assay with the Random_Insertion-hPOMC1-26GLuc clone took advantage of the characteristic of immortal cell lines isolated from neuroblastomas in that they retain the ability to undergo differentiation from the neural crest cells into neural elements with specific phenotypes, depending on the cell line and conditions in which it is held[76], and therefore also retain the ability to exocytose neurosecretory vesicles upon stimulation. Neuroblastoma cell lines have been shown to produce two types of secretory vesicles, including large dense core vesicles[77]. Furthermore, despite the evidence that different stimuli can result in the release of different vesicle types, it has been shown that the exposure to high K^+ levels in neuroblastoma cell lines can stimulate a depolarization strong enough to induce LDCV exocytosis[78]. Supporting this evidence, the Random_Insertion-hPOMC1-26GLuc clone can easily be depolarized in the K^+-HBS buffer[2]. There is a potential, however, that the motor neurons will not respond to stimuli *in vitro* in the same way as the SIMA cells, as discussed in

the section on Membrane fusion (1.2.2). The differentiated cells might only release SVs in response to the K^+-HBS buffer, while the LDCVs might remain in the end terminal until a more appropriate stimulus arrives. One physiological explanation for the delayed release time for LDCVs, is that post-exocytosis, the vesicles must return to the Golgi complex for their membranes to be reused and to be repackaged with their neuropeptide contents[79]. Therefore, it may be that the contents of the LDCVs are reserved for reactions to higher stimuli because of the energy and time required to recycle and repackage their contents. On the other hand, neurotransmitters can be quickly repackaged in SVs directly at the presynaptic terminal[79,80]. Therefore LDCV recycling is quite slow in comparison to that of SVs, which might be detrimental to the application of the LDCV GLuc constructs in the MNs differentiated from the iPSCs.
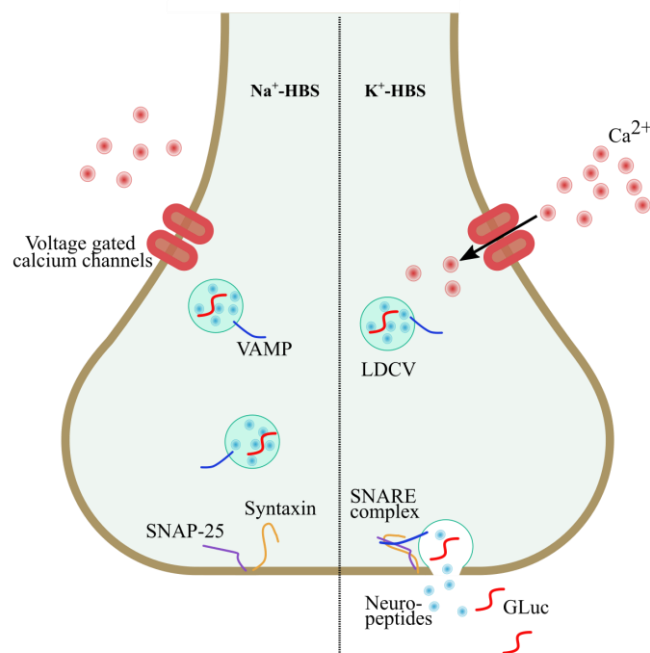
Figure 12: Schematic of luciferase release from presynaptic terminal.

1.6.1.3   Examination of VAMP2-GLuc construct and modifications to release protocol

In order to overcome the potential obstacles presented by the packaging of GLuc in LDCVs, a construct should be designed to produce clones which are able to sort GLuc into SVs. The neuropeptides transported by LDCVs are already associated with the vesicle from the Golgi complex budding, ready for transport to the presynaptic terminal[15,81,82], therefore it is simple to take advantage the sorting domains of the neuropeptides to direct GLuc to the appropriate vesicle. The SV, on the other hand, is not immediately filled with cargo at the Golgi complex, but rather is packaged with neurotransmitters by active transport from the cytosol at

the presynaptic terminal[82]. Therefore the proteins associated with the SV are not being transported in the lumen, but rather are associated with the plasma membrane of the vesicle[83]. For this reason, it is necessary to design the synaptic vesicle construct with GLuc fused to the C-terminus of one of the transmembrane proteins found on the synaptic vesicle. Furthermore, in order to have GLuc associated specifically with the SV, the C-terminus and thus GLuc should be found in the SV lumen. There is only one known SV transmembrane protein fitting this criterion and that is VAMP[84].

The plasmid, which will be created to direct GLuc to the SV, is devised to express a fusion protein of VAMP2 and GLuc. GLuc is fused to the C-terminus of VAMP2, a transmembrane protein that extends into the SV lumen. This construct creates a separate problematic, for which a strategy must be developed to free the fused GLuc from VAMP2. The approach to be tested is to fuse the two proteins together with a linker sequence consisting of the recognition sequence of the TEV (Tobacco Etch Virus) protease[85]. Theoretically, the luciferase release in VAMP2-GLuc cells cannot simply be measured by exposing the cells to depolarization buffer and collecting the cellular supernatant as described for the LDCV system. The GLuc protein is fused to the SV membrane with a segment of sequence coding for a TEV protease recognition element and must therefore be exposed to TEV protease to release the GLuc from the vesicular membrane into the supernatant.

The construct and the steps hypothetically necessary to release GLuc are illustrated in Figure 13. (1) The GLuc protein (red) is fused to the C-terminal of VAMP2 (blue), ensuring its location in the lumen of the synaptic vesicle, which itself is located in the presynaptic terminal of the motor neuron. (2) A stimulus instigates the fusion of the SV to the plasma membrane at the synapse, exposing the GLuc protein and the TEV protease recognition element to the solution surrounding the cell. The TEV protease (purple) in the solution now has access to its recognition element, theoretically enabling the enzyme to cleave GLuc from VAMP2. (3) The TEV protease has cleaved the fusion protein, and GLuc is now freely circulating in the solution surrounding the cell. (4) This solution can be collected and the activity of the free GLuc can be measured (red thunderbolt).

Section 1.6
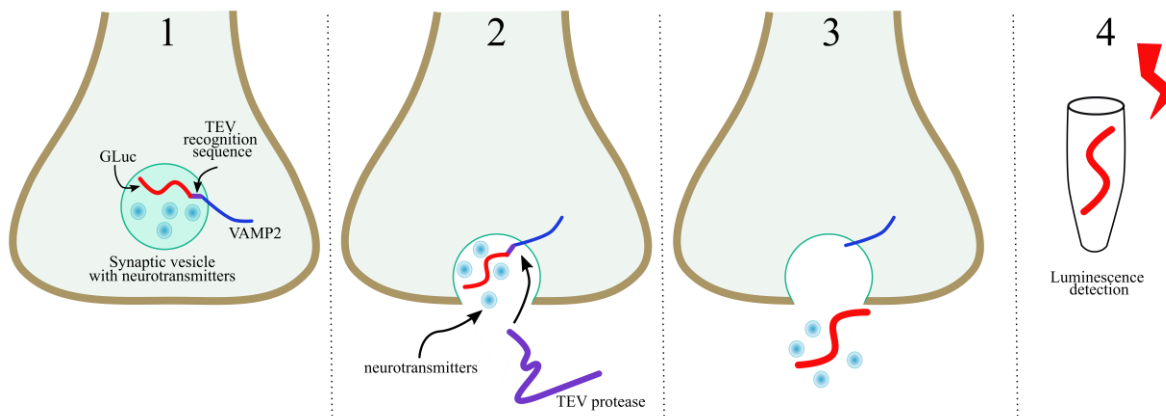
Figure 13: Schematic of the cleavage of GLuc from the VAMP2 fusion protein by the TEV protease upon fusion of the SV with the plasma membrane. Created with Inkscape. (1) SV in the presynaptic terminal with the transmembrane VAMP2 protein (blue) fused by a TEV protease recognition sequence to the GLuc protein (red) in the SV lumen; (2) Fusion event takes place, exposing GLuc (red) to TEV protease (purple); (3) TEV protease has cleaved the fusion protein at its recognition sequence, releasing GLuc into the supernatant; (4) GLuc activity (red thunderbolt) can be measured in supernatant.

The construct directing GLuc to the SV could potentially greatly increase the physiological relevance of the MoN-Light BoNT assay, since the inhibition of neurotransmission from the SV caused by exposure to BoNT could be directly assessed instead of being measured in parallel by GLuc release from the LDCV. Despite this beneficial feature, the construct inherently contains potential obstacles to overcome. First the Gaussia luciferase protein is 19.9 kDa[86], while the VAMP2 protein is only 13 kDa[24], so the fusion of the larger GLuc protein could disrupt the proper sorting of VAMP2. Furthermore, the C-terminus of VAMP2 only contains two amino acids in the vesicular lumen[87], which could make it difficult for the remaining 176 amino acids of the GLuc protein and TEV linker sequence to be fully internalized in the SV lumen. Finally, if the protein is successfully translated and sorted to the SV membrane, for the newly designed luciferase release assay to function correctly, the TEV protease needs to have enough time to access and cut the recognition site upon SV fusion to the plasma membrane.

1.6.2 Explanation of characterization techniques

After the iPS cell line IMR90-4 has been transfected with the double plasmid CRISPR/Cas9 system described above, clones expressing puromycin resistance will be selected and expanded, and the characterization to confirm the precision of the CRISPR editing and the correct GLuc protein localization can begin.

1.6.2.1 Insert confirmation

The first important step in the validation of the CRISPR/Cas9 genetic editing method is to confirm that the donor DNA between the two homology arms in the plasmid has been

correctly incorporated into the AAVS1 safe harbor locus of the transfected cells. This can theoretically be achieved by the amplification by PCR of the region flanking the AAVS1 homology arms and the double strand break point, which can be visualized on an agarose gel[74]. In DNA unaffected by the double strand break, an amplicon of known size corresponding to the WT gDNA should be visible. If the donor DNA is successfully inserted between the HAs, the resulting amplicon should increase by the sequence length of the donor DNA. If the donor DNA has only been incorporated into one allele, both the WT and the insert amplicons should be visible on the gel, whereas an incorporation into both alleles should result in only a single amplicon corresponding to the DNA containing the donor DNA. To be completely certain that this amplicon pertains to the correct genomic locus, the product can be excised from the gel and Sanger sequenced. This method can also be used to exclude any possible mutations incorporated into the sequence during the homology-directed recombination. The amplification region and expected results of the PCR are illustrated in Figure 14.
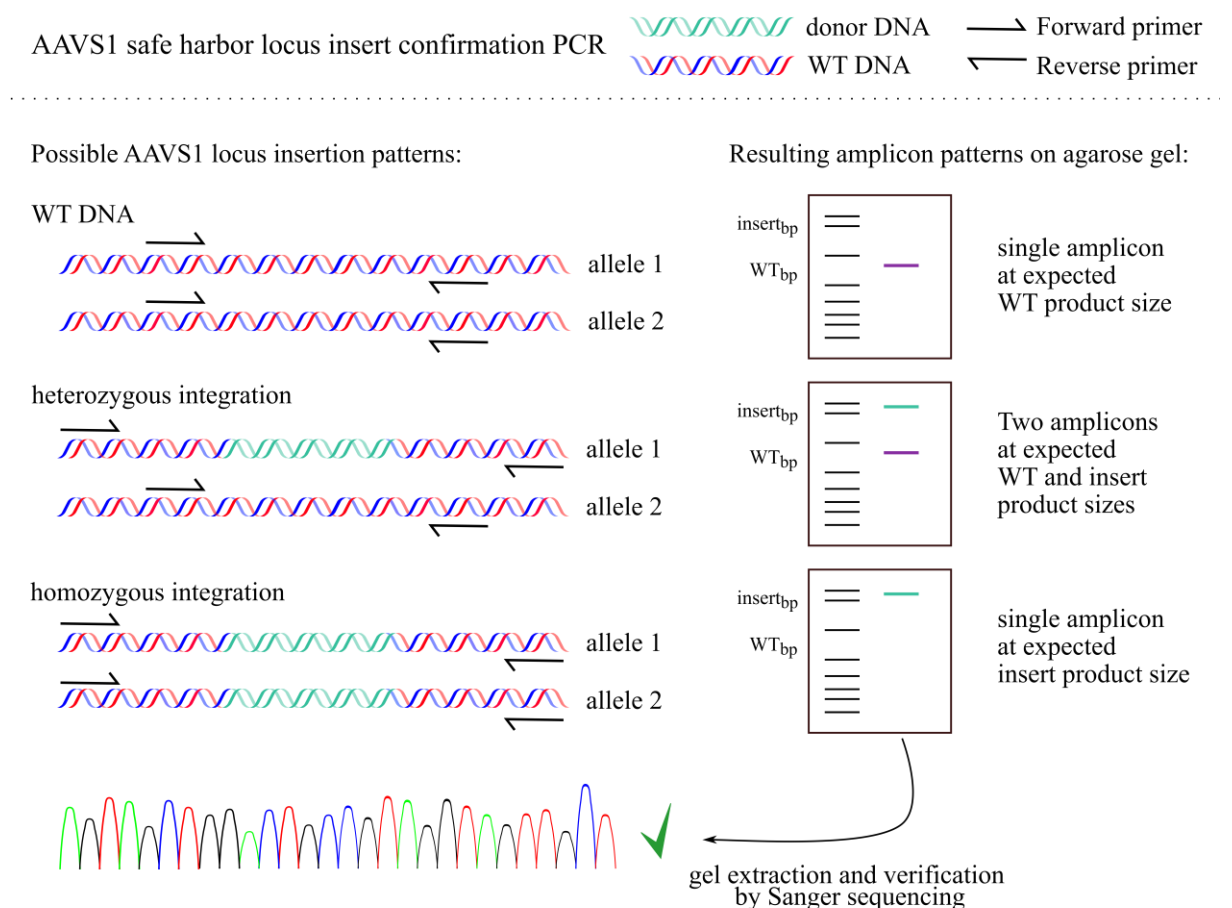
Figure 14: Schematic of insert confirmation PCR of donor DNA at AAVS1 safe harbor locus.

1.6.2.2   Exclusion of off-target integrations

The second important factor in the validation and characterization of the genetically modified clones is to confirm that the integration of the donor DNA into the genome only occurred a single time, at the AAVS1 safe harbor locus, with no off-target events. Several different methods can be useful in the examination of this validation step:

- Southern blot
- Ligation-mediated PCR
- Copy number analysis with qPCR

Southern blot is a standard method to label DNA fragments of various sizes with either radioactive probes or probes labeled with digitonin or streptavidin. The method is carried out by digesting a large quantity of gDNA with a restriction enzyme that tends to cut frequently in the genome. The digested gDNA is run very slowly on an agarose gel to maximize the spread of the DNA fragments, the gDNA is transferred to an appropriate membrane and then hybridized with a probe to detect the DNA sequence of interest[88]. Since the correct insert is known to have occurred at the AAVS1 safe harbor locus, the restriction enzyme recognition sites surrounding this area can be identified and the expected size of the digested gDNA including the donor DNA can be calculated. In the case of a single correct insertion, only a band at this expected size should be visualized by the hybridized probe. In the case of multiple off-target integrations, the correct band should be visible, as well as other bands at various sizes (see Figure 15 for summary of expected results). Southern blot is a traditional, and therefore well-accepted, method to investigate the presence of genetic sequences in a sample[89]. The advantages of the system include its flexibility, whereby the probe hybridization and wash conditions can be easily modified to reinforce the binding and therefore the visualization of the DNA fragment of interest. On the other hand, the method is limited by its lack of sensitivity in the detection of single copies of DNA sequences.

Ligation-mediated PCR is a method that can be used to amplify unknown flanking sequences surrounding DNA that has been randomly inserted into the genome. The method consists of four steps[90] beginning similarly to the Southern blot, with the restriction enzyme digest of the gDNA. The digested DNA fragments are then ligated to adapter oligonucleotides which have been designed to include the complementary overhangs corresponding to the restriction enzyme used for the digest. This ligated DNA library can be used as a template for a PCR reaction in which the primers consist of a known sequence corresponding to the donor DNA and the sequence of the ligated adapter. The amplified fragments can be visualized on an

agarose gel and sequenced to match the flanking genomic sequence to the off-target insertion site in the gDNA (see Figure 15 for summary of expected results). The ligation-mediated PCR helpfully eliminates the need to know the entire sequence of a proposed amplicon. However, the method is prone to yield false negative results because of suboptimal PCR conditions to amplify amplicons of greatly varying size and GC content.

Double control quantitative copy number PCR (dc-qcnPCR) is a new method to quantify elements in the gDNA by normalizing the quantitative amplification of a sequence of unknown copy number to the amplification of a known copy number of autosomal and ChrX genes by qPCR. The sequences of GLuc, the autosomal gene, and the ChrX gene are all amplified from the gDNA of the sample, the gDNA from male and female controls, as well as from a single plasmid which includes all the same sequences. Any differences in primer efficiency or in sample concentrations can be corrected with standardization to the Ct values of the plasmid amplification. Next, the normalized amplification values within each sample in question can be used to calculate the n-fold change in copy number between GLuc, the autosomal gene, and the ChrX gene (Figure 15). While this newly developed technique is initially limited by the necessity to prepare a plasmid with all appropriate control and sequences of interest, once the control plasmid has been prepared the method can be quickly and easily applied to determine the copy number of the sequence of interest. This technique was used to characterize potential off-target integrations of the donor DNA in all clones isolated in this project.
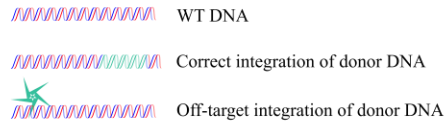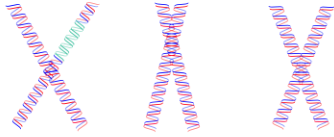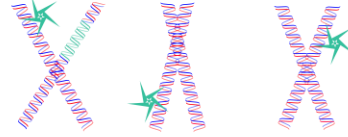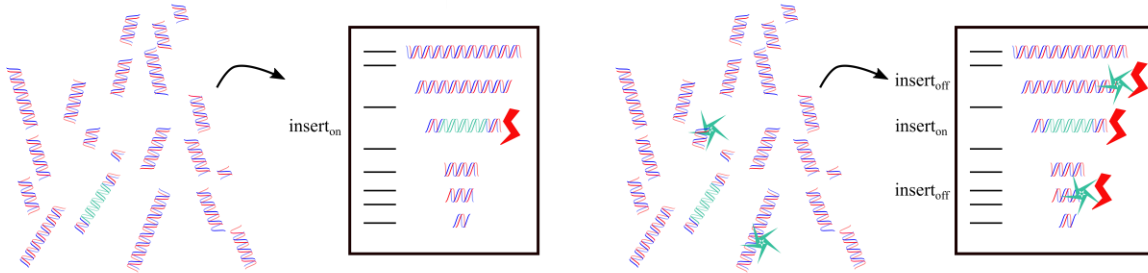
Figure 15: Schematic of methods tested to identify clones with off-target donor DNA integrations. Method 1: Southern blot, digested gDNA is separated by size on an agarose gel and transferred onto a membrane to visualize sequences corresponding to the donor DNA. Method 2: Ligation-mediated PCR, digested gDNA is ligated to oligonucleotide adapters (1) which can be used to amplify (2) unknown DNA segments with a matching primer unique to the donor DNA. These amplicons can be excised from the gel and sequenced to identify the specific on or off target genomic location (3). Method 3: dc-qcnPCR, sequences corresponding to GLuc, an autosomal gene, and an ChrX gene are amplified in gDNA and a control plasmid and copy number of the donor DNA and ChrX gene in male/female DNA is calculated by comparison of Ct values to autosomal control gene.

1.6.2.3 Cellular localization of GLuc protein

The third and most important factor in the validation of the genetically modified clones, after confirming the donor DNA has been integrated in one single locus, is to confirm that the clones are successfully sorting GLuc into secretory vesicles. Two different methods are appropriate to determine the localization of the reporter protein. The first method is differential fractionation[91,92], whereby cells are homogenized to make a solution of cellular components and proteins and then centrifuged at increasing velocities to separate the cellular components into specific subcellular fractions. After the first centrifugation step, the pellet contains nuclear elements, after the second centrifugation step the pellet contains cytosolic elements, and after the last centrifugation step the pellet contains the vesicular/membrane portions of the cell. Each resuspended pellet can be measured for luciferase activity and the localization of luciferase can be deduced (see Figure 16 for illustration of the method). This technique requires a precise degree of homogenization in order to cause the disruption of the cellular membrane, while leaving all other cellular components intact, necessitating an intensive optimization step. The second method is immunofluorescence[93–95], in which the clones are incubated with antibodies that bind to the GLuc, the Golgi apparatus, or the neurosecretory vesicle-associated proteins. These bound antibodies are then visualized through secondary antibodies with fluorescent tags. A series of images can be taken through the cell with a fluorescent microscope and the colocalization of fluorescent voxels, representing the location of specific proteins, can be calculated. Thereby tracking the colocalization of GLuc with the key vesicular sorting pathway organelle, the Golgi apparatus, as well as with other proteins known to localize to LDCVs and SVs (see Figure 16 for illustration of the method). While this method also requires multiple optimization steps in order to determine specific and accurate protein staining, it has the advantage that it allows a more precise visualization and measurement of proteins at a particular timepoint in an intact cell. Immunofluorescence was used to determine the cellular localization of GLuc in the clones prepared for the MoN-Light BoNT assay.

Verification of GLuc localization in clones
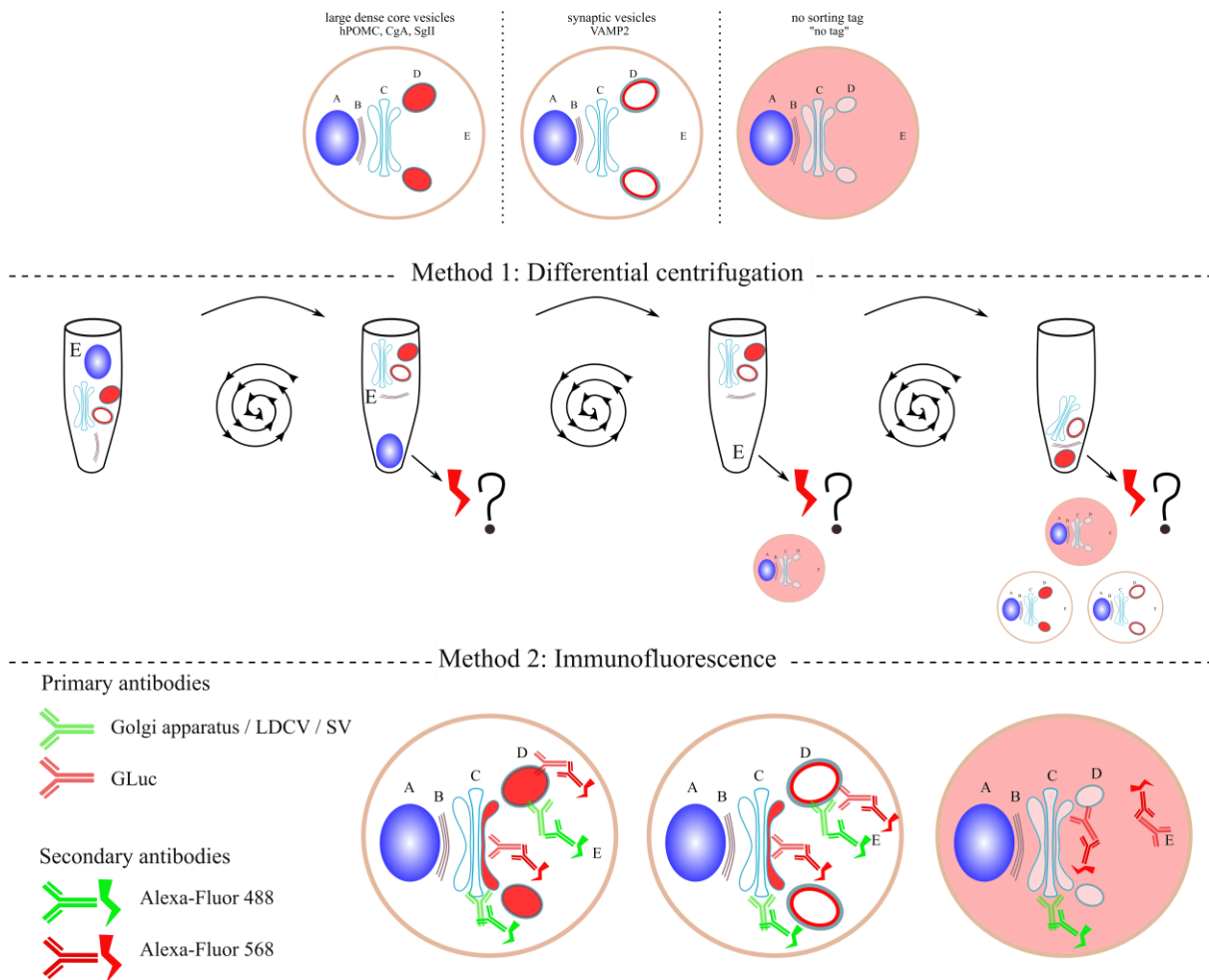
Figure 16: Schematic of methods to verify GLuc protein localization in clones. A = nucleus, B = endoplasmic reticulum, C = Golgi apparatus, D = sprouting vesicle, E = cytosol. Method 1: Differential centrifugation, a series of centrifugation steps separates the nuclear, cytosolic, and membrane fractions. Luciferase activity (red thunderbolt) can be measured to identify cellular location. The no tag GLuc clone is expected to mainly be associated with the cytoplasm, while the remaining clones should all be associated with the final membrane fraction. Method 2: Immunofluorescence, antibodies specific to the proteins of interest (GLuc and Golgi/vesicle markers) are incubated with fixed and permeabilized cells. Secondary antibodies with fluorescent tags bind to the primary antibodies, which can be visualized with a fluorescent microscope.

Once the clones have been characterized and are proven to have passed each checkpoint, which theoretically strengthens the applicability and functionality of this assay in comparison to the Random_Insertion-hPOMC1-26GLuc prototype, they can be included as a possible candidate for the MoN-Light BoNT assay. The clones can be differentiated into motor neurons and the process of optimization of the luciferase release assay itself can begin.

1.6.3 Luciferase release assay description

The basic protocol for the luciferase release assay is relatively simple. The cells are distributed in 96-well plates and allowed to complete their differentiation into the appropriate cell type. Once the cells are mature, they are exposed to either control (Na^+-HBS) or

depolarization (K^+-HBS) buffers. A certain amount of luciferase is spontaneously released into the buffer surrounding the cells, this activity is measured in the control. The depolarization buffer causes a Ca^{2+} influx and therefore the signal for the vesicles to fuse to the presynaptic membrane with the help of the SNARE complex, thereby releasing neurotransmitters and luciferase into the surrounding buffer. The luciferase activity in the depolarization buffer is significantly higher than that in the control buffer (panel A, Figure 17). In order to prove that the increase of luciferase activity after exposure to depolarization buffer is a response to the Ca^{2+} influx and not to irrelevant cellular reactions, the calcium chelator EGTA is added to the control and stimulation buffers, sequestering Ca^{+2} before it can enter the cells[2,96]. The exposure of the cells to the buffers containing EGTA should result in the detection of luciferase activities around the levels, if not lower, of the control buffer (panel B, Figure 17). These first two experiments are done in order to validate that the cell line is worth testing with botulinum neurotoxin, since use of the toxin should be limited to evaluating worthwhile functional assays. Once it has been established that luciferase activity increases upon cellular depolarization, and is inhibited to control levels upon exposure to EGTA, the cells can be tested with BoNT. For the MoN-Light BoNT assay, the cells are exposed to BoNT for 48 hours. During this time, the toxin should have cleaved its specific target, leaving behind dysfunctional SNARE complexes. The cells are subsequently exposed to the control and depolarization buffers. While the depolarization buffer should still provoke an influx of Ca^{2+}, the vesicles are unable to fuse to the presynaptic membrane and luciferase activity should not increase (panel C, Figure 17). Finally, after exposure to a dilution series of BoNT, the changes in luciferase activity can be plotted in a dose response curve and used to determine the LD_{50} equivalent.

Section 1.6
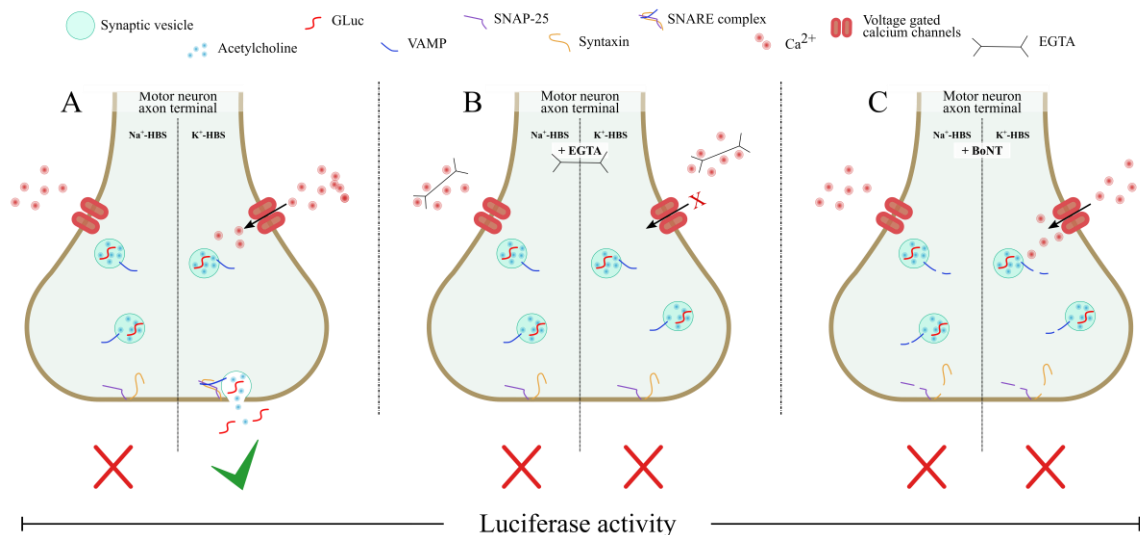
Figure 17: Visual representation of MoN-Light BoNT assay optimization. Under normal conditions (A) exposure to the control buffer (Na$^+$-HBS) does not initiate Ca^{2+} influx and luciferase is not released into the solution surrounding the cell. Exposure to the depolarization buffer (K$^+$-HBS) causes an influx of Ca^{2+} into the presynaptic terminal, initiating the creation of the SNARE complex, and allowing the release of neurotransmitters and luciferase into the surrounding solution, resulting in detection of increased luciferase activity. Upon simultaneous exposure to EGTA, the cell initially reacts the same way to the buffers, but no Ca^{2+} influx can take place because the EGTA has sequestered the supply. Therefore no signal initiates the SNARE complex formation and no vesicle fusion takes place. Luciferase activity does not increase in either buffer. Upon exposure to a high concentration of BoNT (C), the members of the SNARE complex are cleaved (depending on BoNT serotype). The circumstances under the control buffer remain the same, but exposure to the depolarization buffer still instigates Ca^{2+} influx, however the SNARE complex is disrupted and the vesicle cannot fuse to the presynaptic membrane, blocking neurotransmitter and luciferase release. No increase in luciferase activity is detected.

1.7 Summary

Scientists can take advantage of key signals known to occur during gastrulation in order to produce motor neuron populations from iPSCs *in vitro*. Important to neuronal communication are neurotransmitters and neuropeptides, which are packaged in vesicles, and await the appropriate signal to be secreted from the presynaptic terminal. There are two distinct types of neurosecretory vesicles in neurons, the large dense core vesicle and the synaptic vesicle. The production of these vesicles and a portion of their contents occurs through the Golgi apparatus. Important elements packaged into the large dense core vesicles are POMC, CgA, and SgII. Important proteins associated with synaptic vesicles are VAMP, synaptophysin, and VAChT. Once the vesicles are formed, they are transported to the presynaptic terminal to be ready to fuse with the plasma membrane. It appears that each vesicle type initiates the fusion event in response to different signal strengths and durations. Despite this, both vesicles require the SNARE protein complex for fusion with the plasma membrane.

BoNT is a highly toxic substance, which can gain entry into motor neuron presynaptic terminals to cleave specific targets in the SNARE protein family. The cleaved proteins are not able to form a complex and therefore membrane fusion is inhibited. This ends signal

propagation and paralyzes the corresponding muscle fiber. While the gold standard for toxicity assessment of BoNT is the mouse bioassay, the three Rs encourage the development of alternative methods to animal testing. The purpose of this project is to optimize the SIMA Random_Insertion-hPOMC1-26GLuc prototype assay[2], a specific and sensitive method to flexibly determine the toxicity of all serotypes of BoNT in a cell-based *in vitro* assay. The foundation of the assay, to use a Gaussia luciferase (GLuc) reporter directed to neurosecretory vesicles, which can be co-released with neurotransmitters upon cellular depolarization, will remain unchanged. However, the design will be implemented in iPSCs, which can be differentiated into motor neurons, the specific cell target of BoNT. The genetic modification of the cells will take place using CRISPR/Cas9 in order to carry out clean and precise gene editing at the AAVS1 safe harbor locus. The precisely modified, and therefore fully functional motor neuron, should react with the highest sensitivity to exposure of BoNT, strengthening the applicability of this assay for BoNT potency determination.

The SIMA neuroblastoma line and the iPSC line IMR90-4 are genetically modified with CRISPR/Cas. These clones are characterized in order to validate the insertion of the donor DNA in the AAVS1 safe harbor, confirm the single copy insertion of the donor DNA into the gDNA, and verify the correct sorting of GLuc with the neurosecretory pathway (techniques summarized in Figure 18). Post-modification and characterization, the iPSCs are differentiated into motor neurons, the physiological target of BoNTs. The motor neurons are analyzed with qPCR and immunofluorescence to validate the differentiation protocol and to verify that GLuc is still expressed post-differentiation. The motor neurons are utilized in depolarization experiments to test the release of GLuc into the supernatant upon stimulation. If the motor neurons containing either the construct directing GLuc to the LDCV or the SV is able to specifically release the GLuc, the assay might be valid for BoNT toxicity assessment.

CRISPR-modified clones derived from the hPOMC-GLuc and VAMP2-GLuc constructs could be analyzed for all the characterization factors and passed all necessary validation points. Therefore both constructs directing towards LDCVs and SVs could be assessed as motor neurons for their capacity to release GLuc in the luciferase release test. In conclusion, highly innovative and contemporary methods are applied in this project to create the MoN-Light BoNT assay, an *in vivo* simulation assay in cell culture to assess the toxicity of any BoNT serotype. The final objective of this project would be to develop the MoN-Light BoNT assay to the point that its sensitivity to each unique BoNT serotype can reliably and

accurately measure the potency, and its relevance can overtake that of the mouse bioassay, creating a robust and flexible alternative to animal testing.
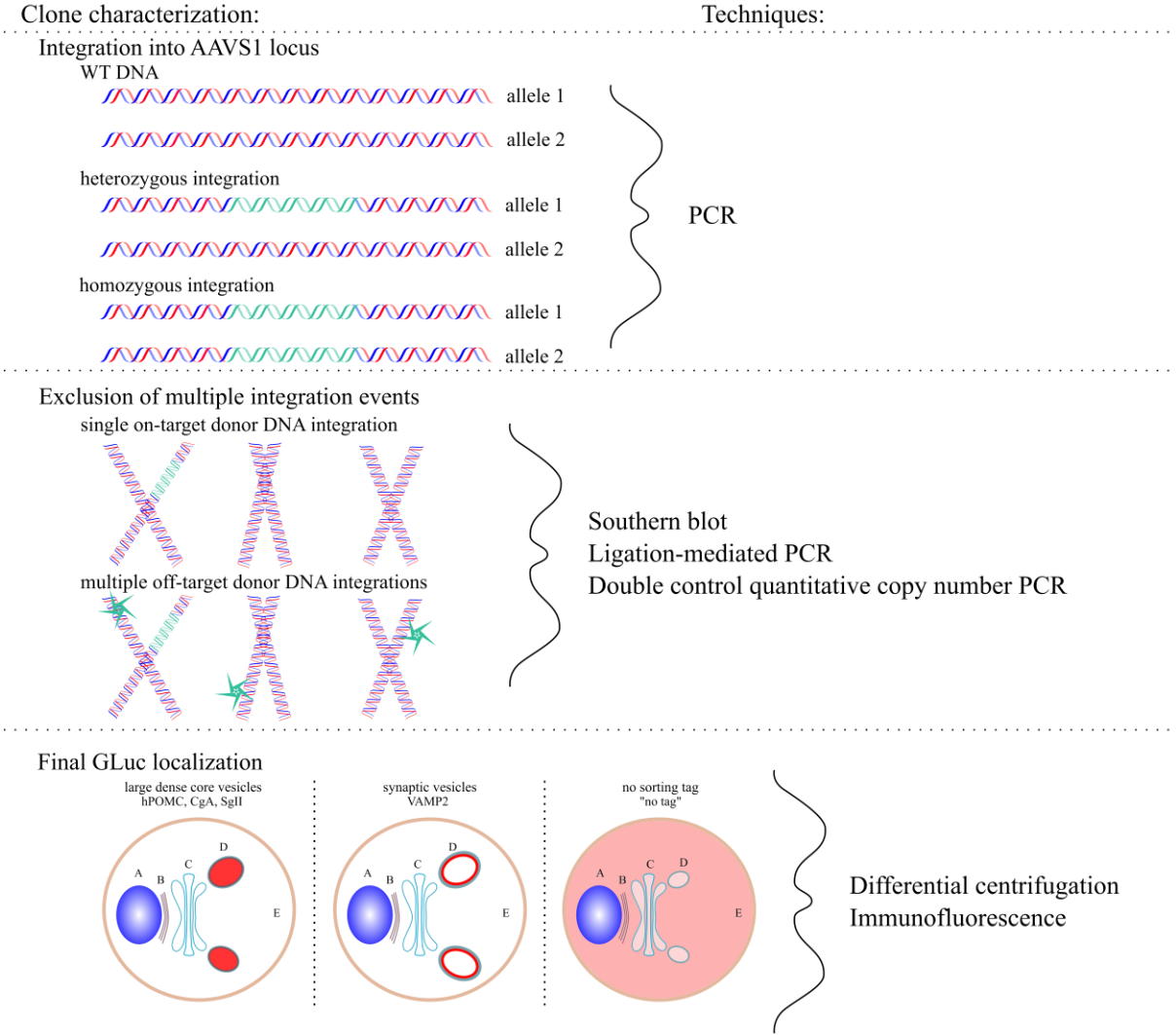
Figure 18: Summary of techniques used to characterize clones.

2 Materials and Methods

2.1 Plasmid production

2.1.1 Standard PCR amplification, digestion and ligation

PCR amplification

All PCR products, unless otherwise specified, were amplified with the PCR master mix described in Table 2 and the thermocycling program described in Table 3. To enable the restriction enzyme to bind and cleave the DNA, all cloning primers included 5' overhangs corresponding to the appropriate restriction enzyme recognition sequence (underlined) and a GCGGCG elongation sequence (bold). PCR primer pairs, as well as their specific annealing temperature and expected product size are found in Table 4. The PCR reaction was run on an agarose gel (2 % agar dissolved in 1x TAE, Table 5) at 130 V for 1 h, visualized with EtBr (0.25 µl/ml), and the desired products were precisely excised from the gel. The amplicon was purified with the Roboclon gel clean-up kit according to the manufacturer's instructions.

Table 2: PCR master mix for amplification of cloning components, final volume of cloning amplification master mix was 50 µl. DMSO and plasmid are placed in braces because they were only added to the reaction if specified in the text.

Master mix	Final concentration
5× GC Phusion buffer	1×
dNTPs (10 mM)	200 µM
Forward primer (10 µM)	0.5 µM
Reverse primer (10 µM)	0.5 µM
{DMSO}	{3 %}
Phusion Polymerase (2 U/µl)	1 U
{plasmid} DNA	10 ng

Table 3: Standard PCR thermal cycling program. Unspecified annealing temperature X is given in Table 4 as "Ann. Temp (°C)" for each specific primer pair.

Step	Temperature (°C)	Duration (sec)	Number of cycles
Denaturation	98	180	1
Denaturation	98	30	
Annealing	X	30	35
Extension	72	60	
Extension	72	600	1

*Table 4: List of primers used for plasmid cloning. F = forward primer, R = reverse primer. Ann. temp. = annealing temperature. *Reverse primer paired with GLuc_SpeI_F. Specific annealing temperature for each primer pair (°C) and the expected product for each amplification (bp).*

Amplicon	Primer name	Sequence (5'-3')	Ann. temp. (°C)	Product size (bp)
Ef1α promoter	Ef1HTLV_PacI_F	**GCGGCG**TTAATTAAGATCTGTAACGGCGCAGAAC	66	1561
	Ef1HTLV_SalI_R	**CGCCGC**GTCGACGTCAGTGGGCAGAGCGCACATC		
Ef1-HTLV promoter	Ef1a_PacI_F	**GCGGCG**TTAATTAATTCACGACACCTGAAATGGAAG	61	499
	Ef1a_SalI_R	**GCGGCG**GTCGACCCCGGGCTGGGCTGAGACCCG		
hPOMC-GLuc	GLuc_SpeI_F	**GCGGCG**ACTAGTCTAGTCACCACCGGCCCCCTTG	72	593
	hPOMC-GLuc_PacI_R	**GCGGCG**TTAATTAACCACCATGCCGAGATCGTGCTG		
no tag GLuc	GLuc_PacI_R*	**GCGGCG**TTAATTAACCACCATGAAGCCCACCGAGAACAACGAAG	72	530
CgA-GLuc	CgA-GLuc_PacI_R*	**GCGGCG**TTAATTAAATGCGCTCCGCCGCTGTC	65	1921
SgII-GLuc	SgII-GLuc_PacI_R*	**GCGGCG**TTAATTAACACCATGGCTGAAGCAAAGACCCA	67	2389
(VAMP) GLuc	GLuc_EcoRI_R*	**GCGGCG**GAATTCCCACCATGAAGCCCACCGAG	72	530
VAMP2	VAMP2_KpnI_F	**GCGGCG**GGTACCAGTGCTGAAGTAAACTATGATG	61	359
	VAMP2_PacI_R	**GCGGCG**TTAATTAACCACCATGTCTGCTACCG		

Table 5: Instructions to prepare 10x TAE buffer.

Component	Concentration
10x Tris-acetate EDTA buffer (TAE)	
Tris-Acetat	137 mM
EDTA	2.7 mM

Digestion

In preparation for cloning, the PCR products were digested with restriction enzymes specific to the cloning step (Table 6, restriction enzymes specified below) using the standard digestion thermocycling program (Table 7). The cloning vector was also digested with the corresponding restriction enzymes and standard digestion protocol (Table 6). It was subsequently dephosphorylated with the addition of 1 µl Fast AP to the digestion reaction and then the sample was incubated further (Table 8). Post-digestion, the PCR products and the

cloning vector were run on an agarose gel (1 % agarose for vectors, 2 % agarose for PCR products) at 130 V for 1 h, visualized with EtBr, and precisely excised from the agarose gel. The products were purified with the Roboclon clean-up kit.

Table 6: Double restriction enzyme DNA digestion reaction.

Reagent	Volume (µl)
2µg plasmid / all gel extracted DNA	20
10× fast digest buffer	2.5
Restriction Enzyme 1	1
Restriction Enzyme 2	1

Table 7: Fast digest thermocycler protocol.

Step	Temperature (°C)	Duration (sec)	Number of cycles
Digestion	37	900	1
Inactivation	70	300	1

Table 8: Fast AP plasmid dephosphorylation cycling program

Step	Temperature (°C)	Duration (sec)	Number of cycles
Dephosphorylation	37	900	1
Inactivation	75	300	1

Ligation

If applicable, the appropriate oligonucleotides were annealed and phosphorylated (Table 9 and Table 10). The dephosphorylated vector and either the gel-purified insert or the annealed oligonucleotides were then ligated together under specific reaction conditions (Table 11 and Table 12). Specific cloning protocols for the key plasmids constructed for this project are described below. The ligation was carried out with 50 ng of plasmid DNA and either a 3 molar excess of the PCR product, or a 1:250 dilution of the annealed and phosphorylated oligonucleotides. The equation to calculate the 3 molar excess of the PCR product is as follows:

$$3 * 50\ ng * (\frac{bp_{insert}}{bp_{plasmid}})$$

Table 9: Annealing and phosphorylation of oligonucleotides

Reagent	Volume (µl)
Oligo-sense (100 µM)	1
Oligo-antisense (100 µM)	1
T4 ligation buffer	1
T4 PNK	0.5
H$_2$O	6.5

Table 10: Thermocycling oligonucleotide annealing protocol

Temperature (°C)	Duration (sec)
37	30 min
95	5 min
to 25	Ramp down at 5 °C/min

Table 11: Ligation reaction of vector and insert.

Reagent	Final concentration
Plasmid (50 ng/µl)	50 ng
PCR product	3 molar excess of plasmid
or	
1:250 dilution annealed and phosphorylated oligonucleotides	4 nM
10× T4 ligation buffer	1×
T4 ligase (5 U/µl)	5 U

Table 12: Thermocycling program for plasmid cloning ligation reaction.

Step	Temperature (°C)	Duration (sec)	Number of cycles
Ligation	22	3600	1
Inactivation	65	600	1

2.1.2 Specific cloning protocols

Plasmid encoding Cas9 endonuclease

The backbone vectors used for the CRISPR/Cas9 co-transfection were provided by Addgene. The first plasmid backbone is the Cas9 endonuclease, which was designed to be expressed from the enhanced specificity Cas9 (eSpCas9(1.1)) plasmid (Figure 19). eSpCas9(1.1) was a gift from Feng Zhang (Addgene plasmid #71814; http://n2t.net/addgene:71814; RRID:Addgene_71814)[72].
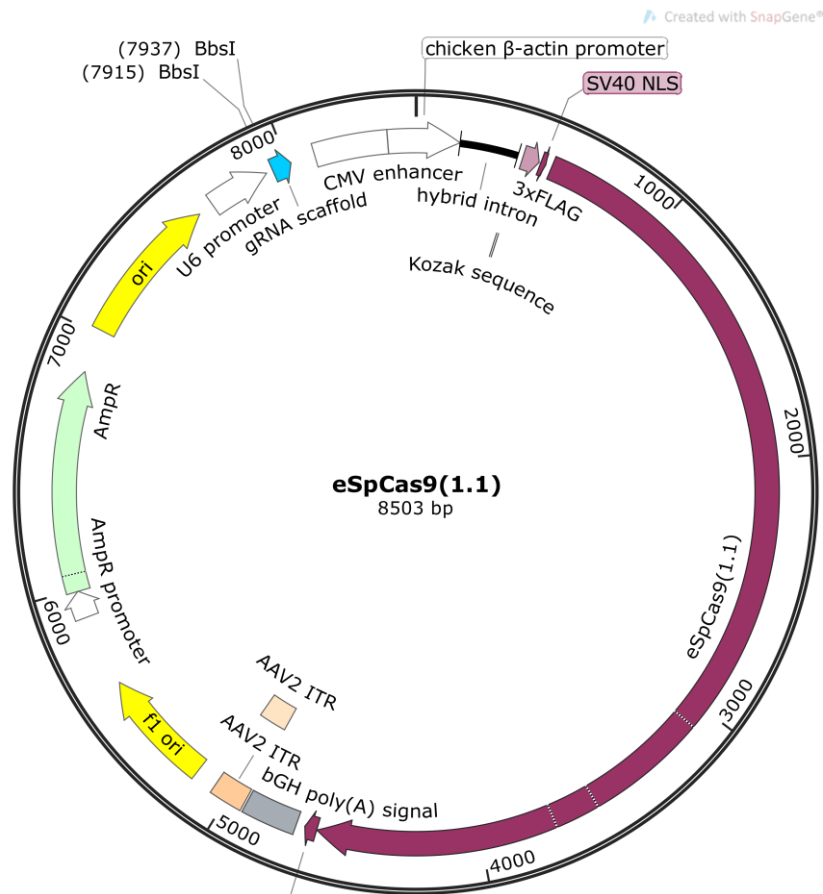
Figure 19: Plasmid map of eSpCas9(1.1).

The sequence of the AAVS1 safe harbor locus T2 sgRNA site[74] was cloned into the eSpCas9(1.1) plasmid at the BbsI cloning site. Two oligonucleotides were designed to contain the sgRNA-T2 recognition sequence and <u>PAM</u> 'CCACTAGGGACAGGAT<u>TGG</u>' and to contain the nucleotide overhangs necessary for sticky-end cloning (Table 13: AAVS1T2_eSpCas9_S and AAVS1T2_eSpCas9_AS). Prior to cloning, the oligonucleotides were annealed and phosphorylated (Table 9 and Table 10). The eSpCas9(1.1) vector was digested with the restriction enzyme BbsI, creating sticky-end overhangs to match the annealed AAVS1-T2 oligonucleotides, and then dephosphorylated according to the standard protocol (Table 6, Table 7, and Table 8). The ligation of the vector and annealed oligonucleotides was also carried out according to the standard protocol (Table 11 and Table 12).

Table 13: Oligonucleotides used for plasmid cloning. S = oligonucleotide in the sense direction 5'-3', AS* = oligonucleotide in the antisense direction 3'-5'

Oligonucleotide name	Sequence
AAVS1T2_eSpCas9_S	CACCGGGGCCACTAGGGACAGGAT
AAVS1T2_eSpCas9_AS*	CCCCGGTGATCCCTGTCCTACAAA
RE_TEV_S	AATTCTCCTTGAAAATATAAGTTTTCCTTAAGAGCGGTAC

RE_TEV_AS*	GAGGAAGTTTTATATTCAAAAGGAATTCTCGC

Plasmid providing AAVS1 safe harbor locus homology arms

The second plasmid backbone used for CRISPR/Cas9 gene editing was pAAVS1-P-MCS. The vector contained sequences for the AAVS1 safe harbor locus homology arms and the puromycin resistance gene (Figure 20). pAAVS1-P-MCS was a gift from Knut Woltjen (Addgene plasmid #80488; http://n2t.net/addgene:80488; RRID:Addgene_80488)[74]. The pAAVS1-P-MCS donor plasmid contains a multiple cloning site (MCS) with the restriction enzyme recognition sites for HincII, AccI, SalI, PacI, SpeI, and SphI between the AAVS1 safe harbor locus homology arms. The donor DNA sequence to be cloned into this vector included the gene expression promoter and tagged or untagged GLuc coding sequences, as described below.
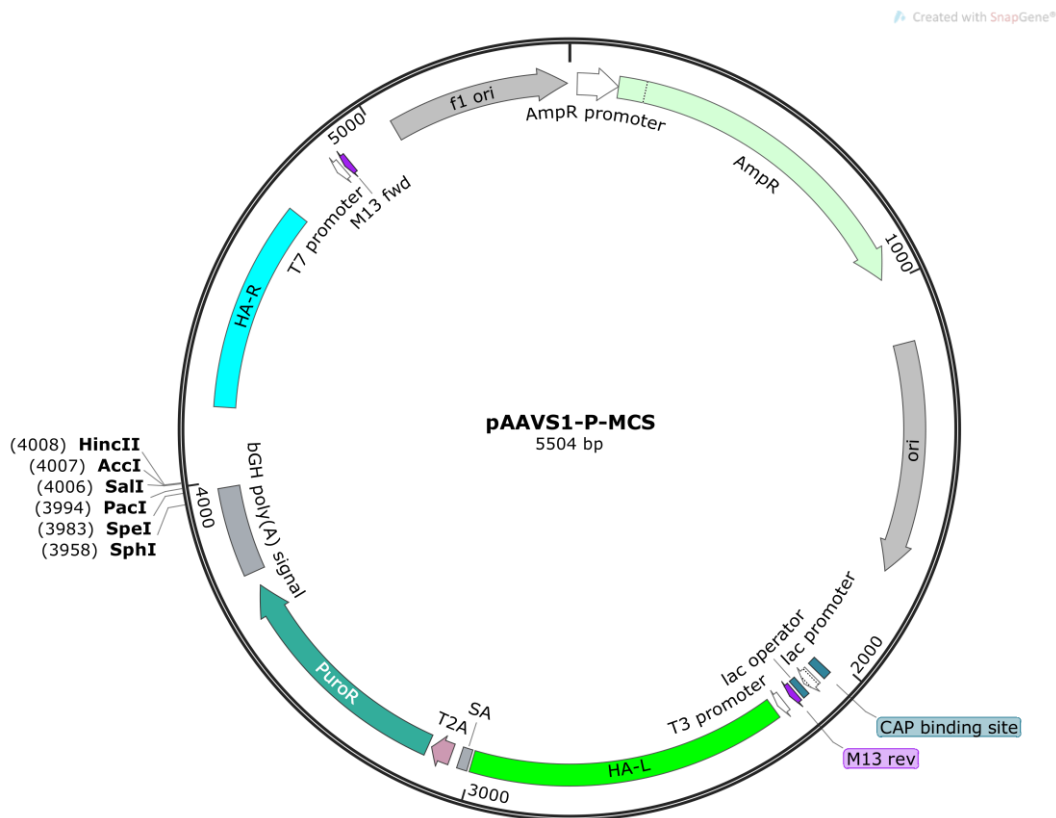
Figure 20: Plasmid map of pAAVS1-P-MCS.

Promoter cloning in pAAVS1-P-MCS

The promoters EF1α or Ef1-HTLV were each cloned into the pAAVS1-P-MCS backbone using flanking primer pairs with 5' overhangs coding for the PacI and SalI restriction enzyme recognition sites (Table 4). The Ef1α promoter was selectively amplified from gDNA

extracted from HEK cells (primer pair: Ef1a_PacI_F, Ef1a_SalI_R). The Ef1-HTLV promoter was selectively amplified from the pNiFty3-SEAP plasmid (also known as pUNO[97], primer pair: Ef1HTLV_PacI_F with Ef1HTLV_SalI_R) using the thermocycling program specific for the promoter amplification (Table 14) and the standard PCR master mix including DMSO (Table 2).

Table 14: Thermocycling program for amplification of Ef1a and Ef1-HTLV promoters. Unspecified annealing temperature X is given in Table 4 as "Ann. Temp (°C)" for each specific primer pair.

Step	Temperature (°C)	Duration (sec)	Number of cycles
Denaturation	98	180	1
Denaturation	98	30	
Annealing	X	20	35
Extension	72	60	
Extension	72	600	1

After PCR amplification, the inserts were run on an agarose gel for 1 h at 130 V, excised from the gel, and purified with the Roboclon gel extraction kit. Both the inserts and the vector plasmid were digested as described in the cloning protocol (Table 6 and Table 7) with the PacI and SalI restriction enzymes, and the vector was dephosphorylated by incubating the reaction with 1 µl Fast AP (Table 8). The digested products were separated on an agarose gel for 1 h at 130 V and then excised and purified. The dephosphorylated and gel-purified vector and the gel-purified insert were then ligated together under the standard reaction conditions (Table 11 and Table 12).

Cloning of GLuc fusion sequences into pAAVS1-P-MCS promoter vectors

All subsequent cloning was carried out with either the promoter vector pAAVS1-P-MCS Ef1α or the promoter vector pAAVS1-P-MCS Ef1-HTLV. The remaining components of the donor DNA to be cloned into the AAVS1-P-MCS backbones were the Gaussia luciferase fusion components henceforth referred to as hPOMC-GLuc, no tag GLuc, CgA-GLuc, SgII-GLuc, and VAMP2-GLuc (see section 1.6.1). The hPOMC-GLuc signal peptide was the sequence encoding amino acids 1-26 of the human proopiomelanocortin protein and was amplified from pcDNA3-hPOMC1-26-GLuc plasmid[2] as a template (Table 4, primer pair: GLuc_SpeI_F and hPOMC-GLuc_PacI_R). The no tag GLuc encoded the Gaussia luciferase protein only, excluding any type of sorting signal, and was amplified from the hPOMC-GLuc plasmid described in section 3.1.2.3 (Table 4, primer pair: GLuc_SpeI_F and GLuc_PacI_R).

The CgA sorting signal contained a section of the 5'UTR and beginning of the *CHGA* genetic sequence and was amplified from pcDNA3-hSCgA-GLuc plasmid[2] as a template (Table 4, primer pair: GLuc_SpeI_F and CgA-GLuc_PacI_R). The SgII sorting sequence contained a section of the 5'UTR and beginning of the *SCG2* genetic sequence and was amplified from pcDNA3-hSgII-GLuc plasmid[2] as a template (Table 4, primer pair: GLuc_SpeI_F and SgII-GLuc_PacI_R). All of the GLuc fusion sequences were amplified with the same forward primer (GLuc_SpeI_F), the reverse primer is specifically labelled according to the tag to be amplified (Table 4), encoding the C-terminus of the protein. The standard PCR master mix and thermocycling program were used to amplify these components (Table 2 and Table 3). The amplified sequences of hPOMC-GLuc, no tag GLuc, CgA-GLuc and SgII GLuc were digested with the restriction enzymes PacI and SpeI (Table 6 and Table 7). The pAAVS1-P-MCS promoter backbones were digested using the standard digestion protocol (Table 6 and Table 7) with the restriction enzymes PacI and SpeI and then dephosphorylated (Table 8). Next, the PCR product and vector were ligated together following the standard ligation protocol (Table 11 and Table 12).

The VAMP2-GLuc donor plasmid is composed of the following three segments: 1) a fusion sequence of the ORF of VAMP2, 2) the TEV protease recognition sequence, and 3) the no tag GLuc sequence. The VAMP2 sequence was amplified from a VAMP2 open reading frame (ORF) plasmid[98] with primers incorporating 5' overhangs corresponding to the restriction enzyme recognition sites for PacI and KpnI (Table 4, primer pair: VAMP2_KpnI_F and VAMP2_PacI_R). Next, the TEV protease recognition sequence segment was constructed with two complementary oligonucleotides, incorporating overhangs corresponding to DNA digested by EcoRI and KpnI (Table 13, oligonucleotide pair: RE_TEV_S and RE_TEV_AS). The oligonucleotides were annealed and phosphorylated according to the standard protocol (Table 9 and Table 10). Then, the no tag GLuc sequence was amplified from the pAAVS1-P-MCS no tag GLuc plasmid (3.1.2.4) with primers incorporating 5' overhangs containing the restriction enzyme recognition sites for SpeI and EcoRI. The no tag GLuc sequence was digested with the restriction enzymes SpeI and EcoRI, and the VAMP2 sequence was digested with the restriction enzymes PacI and KpnI (Table 6 and Table 7). Next, the three segments were ligated together using an overnight ligation protocol (Table 12, ligation step duration extended to overnight at 16 °C). Finally, the entire ligation reaction was amplified with the GLuc_SpeI_F and VAMP2_PacI_R primer pair (Table 4) using the standard PCR master mix and thermocycling program at an annealing temperature of 65 °C (Table 2 and Table 3). The PCR reaction was run

on a gel and the resulting 909 bp product was precisely excised from the gel in order to exclude any potential ligation products which did not contain the short annealed oligonucleotides. This amplified product was then digested with the restriction enzymes SpeI and PacI and then ligated into the pAAVS1-P-MCS Ef1-HTLV plasmid backbone previously digested with the same restriction enzymes (Table 11 and Table 12).

2.1.3 Transformation

To prepare the ligated products for transformation by electroporation into competent E. coli cells, 1 ml SOC medium was warmed to 37 °C. The electroporation cuvette and guide were previously chilled on ice. Working on ice, 1 µl ligation products were added to 50 µl competent E. coli cells, the mixture was quickly added to a cuvette and electroporated at 2.5 kV. Immediately 1 ml warm SOC medium was added to transformed cells and transferred to a 15 ml Falcon tube. Cells were incubated at 37 °C while shaking for 1 h to allow post-electroporation recovery. Next, 100 µl of transformed cells was evenly spread on a LB + 50 µg/ml ampicillin agar plate. The agar plate containing the bacterial transformation was incubated upside down at 37 °C ON. Clones were expanded the next day by picking individual colonies and growing the bacteria in liquid LB medium with 50 µg/ml ampicillin at 37 °C while shaking ON. All solutions used in the transformation process are described in Table 15.

Table 15: Solutions used in transformation.

Component	Concentration
LB medium, pH 7.3	
Yeast extract	0.5 %
Tryptone	1 %
NaCl	1 %
H₂O fill to	1000 ml
LB ampicillin agar plates	
LB Agar	3.2 %
Ampicillin	50 µg/ml
H₂O fill to	1000 ml
SOB medium, pH 7.0	
Tryptone	20 %
Yeast extract	0.5 %
NaCl	10 mM
KCl	2.7 mM
H₂O fill to	1000 ml

Section 2.1

2M Mg^{2+} solution	
MgCl$_2$	20.33 g
MgSO$_4$	12.4 g
H$_2$O fill to	100 ml
SOC medium	
Glucose	20 mM
Mg^{2+}	20 mM
SOB fill to	1 ml

2.1.4 Isolation of plasmid DNA

The clones which were picked and grown in LB medium are resistant to ampicillin and therefore contain an unknown version of the ligated plasmid. The plasmid DNA amplified in the E. coli was isolated with the QIAGEN mini-preparation kit. The isolation was carried out following the manufacturer's instructions, except the plasmid was eluted in 20 µl H$_2$O. Any extra bacteria in LB medium was saved for maxi-preparation in case of positive results. The isolated plasmids were digested with the restriction enzymes used for the vector cloning, to verify if the plasmid contained the correct insert (Table 16). Because the ligation of the 20 bp T2 sequence into the eSpCas9(1.1) hardly changes the size of the 6000 bp plasmid and restriction digest confirmation excising the 20 bp segment is difficult to verify on an agarose gel, the correct ligation was confirmed by PCR amplification to quickly exclude possible clones without the insert. The qPCR was run with the forward sequencing primer eSpCas9_gRNA_F-seq (Table 19) and the AAVS1T2_eSpCas9_AS oligonucleotide (Table 13) using the standard SYBR green master mix (Table 17) and the standard qPCR thermocycling program (Table 18).

Table 16: Reagent mix for restriction enzyme digestion of mini-prepped plasmid for detection of correctly ligated product.

Reagent	Volume (µl)
Mini-prepped plasmid DNA	5
10× Fast digest buffer	2
Restriction Enzyme 1	1
Restriction Enzyme 2	1
H$_2$O	1

Table 17: qPCR master mix to verify the insertion of sgRNA sequence into the eSpCas9(1.1) vector.

Reagent	Final concentration
2× Maxima Sybr Green Master Mix	1x
1:1000 dilution mini-prepped plasmids	1:10,000 dilution
Forward Primer (10 µM)	250 nM

Reverse Primer (10 µM)	250 nM

Table 18: Thermocycling protocol for GLuc expression and gene expression in iPSCs

Step	Temperature (°C)	Duration (sec)	Number of cycles
Denaturation	95	600	1
Denaturation	95	20	
Annealing	60	20	40
Extension	72	20	
Melting curve	95	10	1
	65	5	1
	95		Ramp up from 65 °C

If a plasmid is shown to contain the correctly sized insert, and Sanger sequencing confirmed the correct orientation and sequence of the insert (described below), the remaining transformed bacteria was cultivated in 150 ml LB medium and 50 µg/ml ampicillin ON at 37 °C while shaking. Next, glycerol stocks for the plasmid were prepared with 500 µl transformed E. coli in LB medium and 500 µl glycerol, and stored at - 80 °C for future use. The remaining bacteria were maxi-prepared to extract the plasmid DNA following manufacturer's instructions, except the plasmid was eluted with 1000 µl H$_2$O.

2.2   Sequence verification

All plasmids, as well as genetically modified sequences surrounding the AAVS1 safe harbor locus, were verified by Sanger sequencing with 400-500 ng DNA and 2.5 µM of the relevant primer at Eurofins Genomics. The primers used for amplification or cloning were also those used for sequencing. If the product was too large to provide an accurate and complete sequence in a single sequencing run, such as the insert confirmation product, extra sequencing primers were used to capture the middle portion of the amplicon (Table 19). Resulting chromatograms were aligned to the reference sequence using the SnapGene and Benchling genetic analysis programs (aligned sequences can be found in the Appendix 6.4).

Table 19: Primers used for sequencing of plasmids and AAVS1 safe harbor locus insert confirmation products.

Region	Primer name	Sequence (5' to 3')
sgRNA plasmid confirmation	eSpCas9_gRNA_F-seq	CATGATTCCTTCATATTTGC
	5'ProbeWT_F	TTCAGGTTCCGTCTTCCTCC
	Rpcr-cl-PAC	AGTTCTTGCAGCTCGGTGAC

Insert confirmation (3+ kbp):	GLuc_Rprobe	`ATGAAGCCCACCGAGAACAAC`
	AAVS1_F-seq	`GGTGTCATTCTATTCTGGG`
	HTLV_Probe_R	`GAGGGGCTCGCATCTCTC`
	Rpcr-wt-3'HA	`AGGATCCTCTCTGGCTCCAT`
SgII	SgII_Seq_F	`GAGCCTTGGCAGAACTTTC`
	SgII_Seq_R	`CTTGCCCGAGAGGGATTC`
CgA	CgA_Seq_F	`GATCTCCTTGTAGCCAAGG`
VAMP2	VAMP2_PacI_R	`GCGGCGTTAATTAACCACCATGTCTGCTACCG`
Insert confirmation 3'HA (1 kb)	Insert_EF1_F	`CACGGCGACTACTGCACTTATATACG`
	AAVS1_3'HA_Seq_F	`TGAGATAAGGCCAGTAGCC`
	AAVS1-xHA_R2	`GAGGAGAATCCACCCAAAAGG`

2.3   Cell culture

2.3.1   Standard cell culture maintenance of SIMA, IMR90-4, and HepG2 cell lines

SIMA cells were cultivated in RPMI1640 medium with 1 % penicillin/streptomycin (P/S), 10 % Fetal Calf Serum (FCS), and 2 mM L-alanyl-L-glutamine on untreated cell culture plates. Medium was changed every 2 - 3 days and upon reaching 80 - 90 % confluency, cells were split 1:15 after trypsinization. Undifferentiated IMR90-4 cells were cultured on Matrigel-coated plates in StemMACs™ iPS-Brew XF human cell culture medium containing 1 % P/S in feeder-free conditions. Medium was changed every 2 days and cells were split 1:40 upon reaching 80 – 90 % confluency with 0.5 mM EDTA in PBS (Table 20). Newly split cells were supplemented with 1 µM Y-27632 ROCK inhibitor. Matrigel was applied to the appropriate cell culture plates and incubated at room temperature for at least 2 hours before use. The Matrigel solution was prepared on ice by diluting previously prepared aliquots 1:100 with KnockOut Medium. Cells were frozen for later use in freezing medium. For SIMA cells this freezing medium consisted of 80 % RPMI1640, 10 % FCS, and 10 % DMSO. For IMR90-4 cells this freezing medium consisted of 80 % DMEM/F12, 10 % FCS, 10 % DMSO, and 1 µM Y-27632 ROCK inhibitor. HepG2 cells were cultivated in DMEM with 1 % P/S, 10 % FCS, and 2 mM L-alanyl-L-glutamine on untreated cell culture plates. Medium was changed every 2 - 3 days and upon reaching 80 - 90 % confluency, cells were split 1:15 after trypsinization.

Table 20: Instructions to prepare phosphate buffered saline for cell culture. Autoclave after preparation.

Component	Concentration
Phosphate buffered saline solution (PBS)	
NaCl	137 mM
KCl	2.7 mM
Na_2HPO_4	10 mM

KH$_2$PO$_4$	1.8 mM

2.3.2 Mycoplasma detection test

Cell culture media was screened for mycoplasma. To this end, 50 µl of cell culture medium was collected from the incubated cells. The medium was incubated at 95 °C for 5 min. Treated medium was amplified with the Mycotest primer pair (F: CACCATCTGTCACTCTGTTAACC, R: GGAGCAAACAGGATTAGATACCC) in the PCR reaction described in Table 21. The PCR was run using the thermocycling program found in Table 22.

Table 21: Master mix for mycoplasma detection test PCR.

Reagent	Volume (µl)
Dream-Taq Green Buffer (10×)	5
dNTPs (2mM)	5
Mycotest_F (100 µM)	0.35
Mycotest_R (100 µM)	0.35
Dream Taq Polymerase (5 U/µl)	0.30
Cell culture medium, boiled	2
H$_2$O	37

Table 22: Thermocycling program to test for mycoplasma in cell culture medium.

Temperature (°C)	Duration (sec)	Number of cycles
94	120	1
55	120	1
72	120	1
94	30	
55	60	33
72	60	
72	240	1
4		incubation

2.3.3 Differentiation protocols for SIMA and IMR90-4 cell lines

To differentiate SIMA cells, they were plated on 1× poly-L-lysine (PLL) coated wells and cultured for 72 hours in RPMI1640 medium supplemented with 1× B27 with Vitamin A, 1× N2, 2 mM L-alanyl-L-glutamine, 1 mM non-essential amino acids, and 1 % P/S .

IMR90-4 cells were differentiated into motor neurons using two previously established protocols by Maury *et al*[9] and Du *et al*[8]. Specific technical details on implementing these

50

protocols were kindly provided by Maren Schenke from the University of Veterinary Medicine in Hannover and are summarized in Figure 21. All steps of both differentiation protocols were carried out in N2B27 basis medium consisting of 50 % DMEM/F12, 50 % Neurobasal medium, 1× B27 without Vitamin A, 1× N2, 2 mM L-alanyl-L-glutamine, and 1 % P/S plus the appropriate differentiation reagents described below.

IMR90-4 cell differentiation according to the Du *et al* protocol was carried out as follows: Cells were plated one day before the differentiation was started on Matrigel-coated 6-well plates with 100,000 cells per well. Differentiation began on D0. From D0 - D5 the differentiation medium was composed of N2B27 medium including 100 µM Ascorbic Acid (AA), 3 µM CHIR99021, 2 µM DMH1, and 2 µM SB431542. Medium was changed on D2, D4, and D5. On D6, the cells were split 1:6 with 0.5 mM EDTA (10 min at RT) onto Matrigel-coated 6-well plates. From D6 - D10 the N2B27 medium included 100 µM AA, 1 µM CHIR99021, 2 µM DMH1, 2 µM SB431542, 100 nM Retinoic Acid (RA), and 500 nM Purmorphamine (PMA). Medium was changed on D8 and D10. From D12 - D18 the cells were in suspension culture with N2B27 medium including 100 µM AA, 500 nM RA and 100 nM PMA. Medium was changed on D14 and D16. On D18, the cell suspension was collected and centrifuged for 5 min at 1000 rpm. The cells were resuspended and incubated for 15 min with Accutase at 37 °C to create a single cell suspension. Next, the Accutase was diluted with ten times the volume of DMEM/F12 medium. Then, the cell suspension was centrifuged for 5 min at 1000 rpm and resuspended in N2B27 medium. The cells were plated on Matrigel-coated plates at the following densities: 200,000 cells per well on 6-well plates, 150,000 cells per well on 96-well plates, and up to 150,000 cells per well on coverslips for IF on 24-well plates. The differentiation medium from D18 on consisted of N2B27 with 100 µM AA, 500 nM RA, 100 nM PMA, 100 nM Compound E, 2 ng/ml GDNF, 2 ng/ml BDNF, and 2 ng/ml CTNF. Medium was changed every second day until D30 or until use (summary in Figure 21).

IMR90-4 cell differentiation according to the Maury *et al* protocol was carried out as follows: From D0 - D8 the cells were in suspension. To this purpose, 600,000 cells were prepared per dish and the medium on D0 was composed of N2B27 with 500 nM AA, 3 µM CHIR99021, 2 µM DMH1, 2 µM SB431542, and 5 µM Y-27632 ROCK inhibitor. D2 medium consisted of N2B27 with 500 nM AA, 3 µM CHIR99021, 2 µM DMH1, 2 µM SB431542, 100 nM RA, and 500 nM SAG. The N2B27 medium on D4 though D8 was supplemented with 500 nM AA, 100 nM RA and 500 nM SAG. Medium was changed on D6 and D8. On D9, the cell suspension was collected and centrifuged for 5 min at 1000 rpm. The cells were

resuspended and incubated for 15 min at 37 °C in Accutase to make a single cell suspension. The Accutase cell suspension was diluted with ten times the volume of DMEM/F12 medium. Then, the cell suspension was centrifuged for 5 min at 1000 rpm and resuspended in N2B27 medium. Finally cells were plated at the following densities: 300,000 cells per well on 6-well plates, 150,000 cells per well on 96-well plates, and 100,000 cells per well on coverslips for IF on 24-well plates. The medium on D9 contained N2B27 with 10 µM DAPT. The medium on D11 was composed of N2B27 with 15 µM AA, 10 µM DAPT, 5 ng/ml GDNF, 5 ng/ml BDNF, 5 ng/ml CNTF, and 1 µg/ml dbcAMP. The next medium change occurred on D14 and all subsequent medium changes took place every two days with N2B27 medium containing 15 µM AA, 5 ng/ml GDNF, 5 ng/ml BDNF, 5 ng/ml CNTF, and 1 µg/ml dbcAMP. After D30, the cells were analyzed as candidate mature motor neurons (summary in Figure 21).
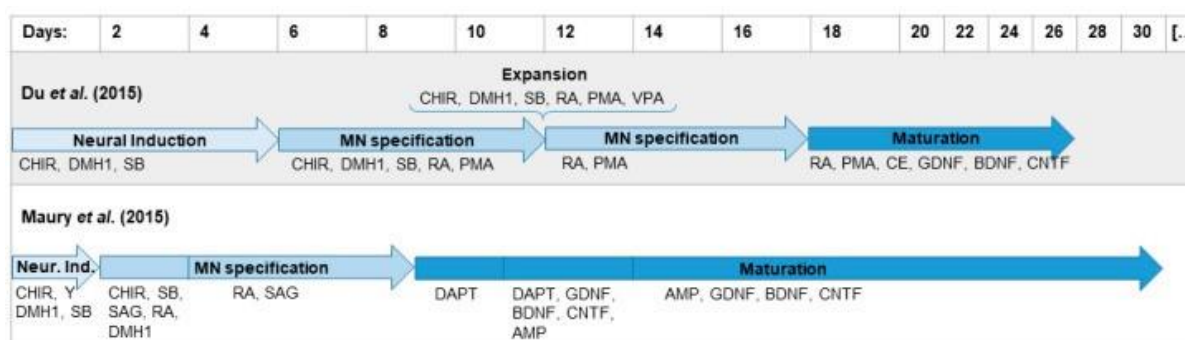
Figure 21: Schematic of protocols used to differentiate iPSCs into MNs. Adapted figure published in Schenke et al[99], licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/)

2.4 Transfection and clone selection

2.4.1 SIMA transfection

Transient transfection

The SIMA cells were transiently transfected in order to verify the expression of luciferase and therefore the functionality of each plasmid. The cells were seeded with 50,000 cells in a single well of a 6-well plate and transfected the following day with 2 µg total plasmid in TurboFect transfection medium. All transfections were performed with two plasmids: the eSpCas9(1.1)+T2sgRNA plasmid and the GLuc donor plasmid derived from AAVS1-P-MCS in a ratio of 1:4. Accordingly, the transient transfection contained 0.5 µg eSpCas9(1.1)+T2sgRNA plasmid and 1.5 µg pAAVS1-P-MCS_X-GLuc plasmid, whereby "X" refers to the specific GLuc construct used in the transfection. To begin the transfection the culture medium was changed to serum-free RPMI1640 including the reagents described by the

52

manufacturer (Table 23). The cells were incubated at 37 °C. Transgene expression was analyzed by measuring luminescence in cell lysate after 24 - 48 hours.

Table 23: SIMA transient transfection components. X = specific AAVS1-P-MCS GLuc construct.

Reagent	Quantity per well
Serum-free RPMI1640 medium	Fill to 200 µl
eSpCas9(1.1)+T2sgRNA	0.5 µg
pAAVS1-P-MCS_X	1.5 µg
TurboFect Transfection Reagent	4 µl

Stable transfection

The stable transfection of SIMA cells was carried out according to the TurboFect transfection protocol with 10 µg total plasmid DNA in one well of a 6-well plate. The cells were seeded at 50,000 cells per well and transfected the following day with TurboFect transfection medium (Table 24). The transfection medium was added to the adherent cells and incubated for 48 h at 37 °C. To isolate the cells harboring stably integrated donor DNA, the medium was changed to the standard care RPMI1640 medium plus 1 µg/ml puromycin antibiotic. The donor DNA contains the sequence for the puromycin resistance gene and those cells expressing this gene are likely also to contain the GLuc fusion sequence of interest. The RPMI1640 medium including puromycin was changed every two days. The medium in which the cells had been incubating, containing primarily dead cells, was transferred to a clean well in a 6-well plate to screen for any remaining living cells. Medium change continued until clusters of healthy cells began to grow after approximately 2 - 4 weeks. These healthy cell clusters were trypsinized and transferred to a well of a 96-well plate as a single cell dilution. Once monoclonal cell clusters began to grow they were trypsinized and expanded into 6-well plates for propagation, freezing, and cellular analysis.

Table 24: SIMA stable transfection components. X-GLuc = specific AAVS1-P-MCS GLuc construct

Reagent	Quantity per well
Serum-free RPMI1640 medium	Fill to 1000 µl
eSpCas9(1.1)+T2sgRNA	2.5 µg
pAAVS1-P-MCS_X-GLuc	7.5 µg
TurboFect Transfection Reagent	20 µl

2.4.2 IMR90-4 transfection

The IMR90-4 transfection was carried out according to the Lipofectamine 300 transfection protocol. Cells were seeded in two wells of a 6-well plate and the transfection was

begun once the cells reached 60 – 70 % confluency. The plasmids were added in a 1:4 ratio of eSpCas9(1.1)+T2sgRNA plasmid to the appropriate GLuc donor plasmid (pAAVS1-P-MCS_X-GLuc). The two transfection master mixes (Table 25), master mix 1 and master mix 2 were combined, incubated at RT for 15 min and then added to cells containing fresh standard iPSC-Brew medium. The cells and transfection reagents were incubated at 37 °C for 48 h. The cells in one well were removed and lysed to verify transient luciferase expression. The cells in the second well were selected for the stable genomic integration of donor DNA by removing the transfection medium and applying standard iPSC-Brew medium containing 500 ng/ml puromycin. The selection medium was changed every two days. Once most cells had died and only a few healthy clusters of cells remained, these healthy clusters were split into 96-well plates for single cell dilution and monoclonal isolation. Cells on the 96-well plate were incubated in CloneR medium at the concentrations specified by the manufacturer. The CloneR reagent was diluted in the standard iPSC-Brew medium plus 500 ng/ml puromycin instead of in mTeSR medium, as normally recommended by the manufacturer. CloneR supports the survival of single cells and was observed to inhibit spontaneous differentiation of the isolated clones. Once a cluster of healthy cells had grown in the well of the 96-well plate, these cells were transferred to a 24-well plate and then a 6-well plate for propagation, freezing, and cellular analysis.

Table 25: Transfection components for IMR90-4 cells with Lipofectamine 3000 kit. X-GLuc refers to the specific tagged GLuc applied in the transfection reaction.

Reagent	Quantity per well
Master mix 1	
Opti-MEM medium	125 µl
Lipofectamine 3000 reagent	7.5 µl
Master mix 2	
Opti-MEM medium	125 µl
eSpCas9(1.1)+T2sgRNA	625 ng
pAAVS1-P-MCS_X-Gluc	1875 ng
P3000 reagent	5 µl

2.5   Luciferase activity measurement

Synthetic coelenterazine was resuspended in EtOH to a concentration of 2 mM stock solution. Two h before luminescence measurement, the 2 mM stock was diluted to 10 µM in H_2O and stored in the dark until use. Lysates were used to confirm luciferase expression in monoclonal cell lines before expansion or to measure background luciferase activity in the

54

luciferase release assay (described in detail below, section 2.15). To confirm GLuc expression, lysates were prepared by harvesting a clone at 50 - 70 % confluency, isolating 20 % of the cells, and incubating the cells in 20 µl 1× passive lysis buffer for 10 min on ice. All luminescence measurements were made in a FLUOstar Omega microplate reader. Luciferase activity was measured in 20 µl lysate in a white-walled clear-bottomed 96-well plate by automatically injecting 100 µl of a 10 µM coelenterazine solution to the sample.

For the luciferase release assay, the background lysate luciferase activity was measured by injecting 100 µl of a 10 µM solution of coelenterazine directly onto the cells cultured in a white-walled clear flat-bottomed 96-well plate. Luciferase activity in the supernatant was measured by adding 20 µl of the sample to a white-walled white-bottomed 96-well plate and then automatically injecting 100 µl of a 10 µM solution of coelenterazine to each well.

2.6   DNA extraction

DNA extraction was carried out using the QIAGEN DNeasy Blood and Tissue extraction kit, according to the protocol with the following two modifications. First, after the second washing step, the empty column was centrifuged at 14,000 rpm to remove any residual wash buffer. Second, when preparing gDNA for Southern blotting, the DNA was eluted in two steps, first in 50 µl H_2O, and then a second elution in 150 µl H_2O. The first elution yielded a more highly concentrated gDNA sample, as required for Southern blotting. The gDNA from the second elution was diluted to the necessary concentration and used for qPCR.

2.7   Ethanol precipitation of DNA

Samples with low gDNA concentrations following DNA extraction were subjected to ethanol precipitation. To this end, sodium acetate (Table 26) was added to the DNA sample to a final concentration of 300 mM. The sample was vortexed and then 2.5 – 3 volumes of 95 % EtOH were added. The sample was incubated on ice for 15 min, centrifuged at 4 °C for 30 min at 14,000 rpm, and then the supernatant was carefully discarded. Next, the pellet was rinsed with 70 % EtOH and then centrifuged at 4 °C for 15 min at 14,000 rpm. Finally, the supernatant was discarded and sample was resuspended in 20 µl dH_2O.

Table 26: Instructions to prepare 3M sodium acetate for the ethanol precipitation of gDNA.

3M sodium acetate	
CH₃COONa	24.61g
Acetic acid adjust to	pH 5.2
H₂O fill to	100 ml

2.8   Insert confirmation (PCR)

The protocol to test the insert confirmation at AAVS1 safe harbor locus is based on the polymerase chain reaction. To establish the primary insert confirmation, a pair of primers were designed to flank the locus at which the Cas9 endonuclease induces the double strand break in the gDNA. Multiple primer pairs were tested, as described in section 3.2.1 (Table 27). Each primer pair was tested in a 20 µl reaction (PCR master mix found in Table 28, temperature gradient thermocycling program found in Table 29). Finally, the insert confirmation was successfully carried out using the primers 5'probeWT_F and Rpcr-wt-3'HA at a 70 °C annealing temperature (Table 27). All possible PCR products were separated on an agarose gel and visualized with EtBr to confirm the presence of WT and/or insert bands. The primers used for the secondary confirmation PCR of the 3'HA were AAVS1_3'HA_F and AAVS1-xHA_R2. Amplification was carried out with the same master mix and thermocycling program found in Table 28 and Table 29. The products of both the primary and secondary PCR sets were amplified and separated on an agarose gel and the amplicons were excised and gel-purified, as described previously, for Sanger sequencing. This step confirms the amplification of the correct product, as well as excludes any mutations that might have been incorporated into the gDNA during the homology-driven recombination.

Table 27: Primers used to test insert confirmation. Annealing temperature is included for the optimized final primer pairs.

Primer name	Sequence (5'-3')	Annealing temperature (°C, if amplified)
Fpcr-803*	TCGACTTCCCCTCTTCCGATG	-
Rpcr-cl-804*	GAGCCTAGGGCCGGGATTCTC	-
Rpcr-wt-183*	CTCAGGTTCTGGGAGAGGGTAG	-
Fpcr_5'ofHA	TTGCTCTCTGCTGTGTTGCT	-
5'probeWT_F	TTCAGGTTCCGTCTTCCTCC	70
Rpcr-wt-3'HA	AGGATCCTCTCTGGCTCCAT	
AAVS1-xHA_R4	AGGGGAACGGGGATGCAG	-
AAVS1-xHA_R3	GCTCAGTCTGAAGAGCAGAGC	-
AAVS1_3'HA_F	AGCCAGTACACGACATCACT	70
AAVS1-xHA_R2	GAGGAGAATCCACCCAAAAGG	

Table 28: Temperature gradient PCR master mix to determine appropriate annealing temperature for PCR products.

Reagent	Final concentration
5× GC buffer	1×

dNTPs (10 mM)	200 µM
5'probeWT_F (10 µM)	0.5 µM
Rpcr-wt-3'HA (10 µM)	0.5 µM
DMSO	3 %
Phusion Polymerase (2 U/µl)	0.4 U
DNA	10-20 ng

Table 29: Temperature gradient PCR thermocycling protocol for insert confirmation.

Step	Temperature (°C)	Duration (sec)	Number of cycles
Denaturation	98	30	1
Denaturation	98	45	
Annealing	55-70	45	35
Extension	72	120	
Extension	72	600	1

The CgA and SgII sorting signal sequences are both much longer than the hPOMC sequence, therefore the expected product size of the donor DNA amplicon was much larger. Therefore, the protocol for the amplification of CgA-GLuc and SgII-GLuc donor DNA segments incorporated into the IMR90-4 gDNA was optimized (see Table 30 for all expected PCR product sizes with primer pair 5'probeWT_F and Rpcr-wt-3'HA, and the 3'HA amplification). The optimized PCR master mix description is found in Table 31 and the adjusted thermocycling program is found in Table 32.

Table 30: Expected product sizes for each insertion confirmation product. The WT AAVS1 safe harbor locus, hPOMC-GLuc, CgA-GLuc, SgII-GLuc, and VAMP2-GLuc inserts were amplified with the primer pair 5'probeWT_F and Rpcr-wt-3'HA. The 3'HA amplification product was amplified with the primer pair AAVS1_3'HA_F and AAVS1-xHA_R2.

Name of amplicon	Product size (bp)
WT AAVS1 safe harbor locus	856
hPOMC-GLuc insert	2950
CgA-GLuc insert	4289
SgII-GLuc insert	4757
VAMP2-GLuc insert	3292
3'HA amplification	1000

Table 31: Modified master mix formulation for insert confirmation of CgA-GLuc and SgII-GLuc donor DNA segments incorporated in AAVS1 safe harbor locus.

Reagent	Final concentration
5× GC buffer	1×
dNTPs (10 mM)	200 µM
5'probeWT_F (10 µM)	0.5 µM
Rpcr-wt-3'HA (10 µM)	0.5 µM

DMSO	9 %
Phusion Polymerase (2 U/µl)	0.4 U
DNA	10-20 ng

Table 32: Modified thermocycling program for insert confirmation of CgA-GLuc and SgII-GLuc donor DNA segments incorporated in AAVS1 safe harbor locus.

Step	Temperature (°C)	Duration (sec)	Number of cycles
Denaturation	98	30	1
Denaturation	98	10	
Annealing	67.5	30	35
Extension	72	150	
Extension	72	600	1

2.9   Southern Blot

Southern blot was performed according to the DIG Application Manual for Filter Hybridization supplied by Roche. PCR preparation of southern blot probes was carried out using the Roche DIG PCR kit (Table 33). Annealing temperatures were determined by temperature gradient PCR indicated by "X" in the DIG probe thermal cycling protocol (Table 34). The final list of all probes tested with their optimal annealing temperature is found in Table 35.

Table 33: PCR master mix for Southern blot DIG probes.

Final concentration DIG-labeled product	Reagents	Final concentration non-DIG control
1×	**PCR buffer with MgCl₂, 10× conc. (vial 3)**	1×
1×	**PCR DIG mix, 10× conc. (vial 2)**	0
0	**dNTP stock solution, 10× conc. (vial 4)**	1×
1 µM	**Forward primer (10 µM)**	1 µM
1 µM	**Reverse primer (10 µM)**	1 µM
1×	**Enzyme mix, Expand High Fidelity (vial 1)**	1×
40 ng	**gDNA**	40 ng
fill to 50 µl	**H₂O**	fill to 50 µl

Table 34: Thermal cycling conditions for DIG probe PCR.

Step	Temperature (°C)	Duration (sec)	Number of cycles
Denaturation	95	120	1
Denaturation	95	30	30
Annealing	X	30	

Extension	72	150	
Extension	72	420	1

Table 35: List of Southern blot probe sequences and annealing temperatures for PCR amplification.

Probe name	Primer name	Primer sequence	Annealing temperature (°C)
WT probe	5'probeWT-F	TTCAGGTTCCGTCTTCCTCC	58
	5'probeWT-R	CGGGTTGGAGGAAGAAGACT	
Efa1HTLV	Efa1HTLV_ProbeF	GTAACGGCGCAGAACAGAA	52
	Efa1HTLV_ProbeR	AACCGGTGCCTAGAGAAGGT	
HTLV	Efa1HTLV_ProbeF	GTAACGGCGCAGAACAGAA	52
	HTLV_Probe_R	CTGAAGCTTCGAGGGGCTC	
Puromycin	Puro_ProbeF	GTCACCGAGCTGCAAGAACT	58
	Puro_ProbeR	GCTCGTAGAAGGGGAGGTTG	
GLuc-F1R1	GLuc_Fprobe	CTAGTCACCACCGGCCCC	58
	GLuc_Rprobe	ATGAAGCCCACCGAGAACAAC	
GLuc-F2R2	GLuc_Fprobe2	GTCAGAACACTGCACGTTGG	60
	GLuc_Rprobe2	AAGACTTCAACATCGTGGCC	
GLuc-F3R3	GLuc_Fprobe3	CCCCTTGATCTTGTCCACCT	60
	GLuc_Rprobe3	GGTCGATCTGTGTGTGGACT	
GLuc-F3R4	GLuc_Fprobe3	CCCCTTGATCTTGTCCACCT	60
	GLuc_Rprobe4	GCACGCCCAAGATGAAGAAG	
GLuc-F2R4	GLuc_Fprobe2	GTCAGAACACTGCACGTTGG	60
	GLuc_Rprobe4	GCACGCCCAAGATGAAGAAG	

On Day 1, 15 µg DNA was digested with 5 U restriction enzyme per µg gDNA for 18 h ON at 37 °C. On Day 2, the digested samples were separated ON at 4 °C at 20 V on an agarose gel. On Day 3, the agarose gel was placed upside down (wells facing down) into an appropriately sized container. The depurination, denaturation, and neutralization steps were all carried out according to the manufacturer's instructions while gently shaking (Table 36).

Table 36: Non-ready-made solutions for Southern blot

Depurination solution	
HCl	250 mM
Denaturation solution	
NaOH	0.5 M
NaCl	1.5 M
Neutralization solution	
Tris-HCl (pH 7.5)	0.5 M
NaCl	1.5 M

20× SSC solution (pH 7.0)	
NaCl	3 M
Sodium citrate	300 mM
Low stringency washing buffer	
SSC solution	2×
SDS	0.1 %
High stringency washing buffer	
SSC solution	0.1×
SDS	0.1 %
Stripping buffer	
NaOH	0.2 M
SDS	0.1 %

Figure 22: Capillary transfer system from agarose gel to nylon membrane

The capillary transfer system was assembled according to the manufacturer's instructions (Figure 22) and transfer proceeded for 16 h at RT. Throughout the subsequent pre-hybridization, hybridization, and washing steps, the membrane kept damp. The blot was placed into a hybridization tube with 10 ml pre-hybridization buffer per 100 cm^2 for at least 1 h at the specific hybridization temperature (T_{hyb}). The optimal hybridization temperature range was calculated using the following formulae, whereby T_m = melting temperature:

$$T_m = 49.82 + \left(0.41 * \%GC_{probe}\right) - \left(\frac{600bp}{bp_{probe}}\right)$$

$$T_{hyb} = T_m - (20 \text{ } to \text{ } 25) = optimal \text{ } hybridization \text{ } range$$

Table 37: Probe size and melting and hybridization temperatures for all Southern blot probes tested in this project

Probe	Size (bp)	T_m (°C)	T_{hyb} (°C)
GLuc F1+R1	510	72.8	53 - 48
GLuc F2+R2	395	72.3	52 - 47
GLuc F3+R3	145	71.8	49 - 44
GLuc F3+R4	299	71.7	52 - 47

60

GLuc F2+R4	221	70.7	56 - 51
Puro	382	78.2	58 - 53
Ef1a-HTLV	430	73.0	53 - 48
HTLV	266	73.2	53 - 48

The pre-hybridization, hybridization, and post-hybridization steps were carried out with ready-made supplies from the DIG Wash and Block Buffer Set from Roche and the stringency buffers (Table 36) according to the manufacturer's instructions. The pre-hybridization and hybridization were carried out within the temperature range calculated for each probe (Table 37). After the washing steps, the membrane was placed DNA-side up between two plastic sheet protectors and 1 ml chemiluminescent substrate (0.25 mM CSPD) was added dropwise to the blot. Following exclusion of air bubbles from the bag, the membrane was incubated for 10 min at 37 °C. Finally, luminescence was measured every 30 sec for 45 min.

To strip the blot for future reuse, the membrane was thoroughly rinsed in dH$_2$O. The membrane was incubated twice in stripping buffer for 15 min at 37 °C and then washed in 2× SSC for 5 min at RT. The hybridization and detection were either repeated at a different temperature or the membrane was stored in 2× SSC at 4 °C.

2.10 Ligation-mediated PCR

2.10.1 Preparation of gDNA for ligation-mediated PCR

The ligation-mediated PCR protocol was adapted from the method reported by O'Malley *et al*, 2007[100]. A total of 2.5 µg of gDNA of the SIMA hPOMC-GLuc clone 4 was digested overnight with either HindIII, BspHI, or AseI (Table 38) at 37 °C. The digested gDNA was purified and concentrated by ethanol precipitation using the protocol described in section 2.7 and resuspended in 20 µl H$_2$O. The overhang adapters were annealed and phosphorylated using the standard protocols (Table 9 and Table 10). The antisense oligonucleotide sequence was complementary to all three sense oligonucleotides, each lending a unique overhang designed to match a DNA sequence a specific restriction recognition site (Table 39). The digested gDNA and the compatible adapter oligonucleotides were ligated using the standard ligation master mix (Table 11), but the ligation time was conducted overnight at 16 °C.

Table 38: Digestion components of the ligation-mediated PCR. Restriction enzyme X was either HindIII, BspHI, or AseI.

Reagent	Quantity
gDNA	2.5 µg
Restriction enzyme X	5 µl
10× fast green buffer	10 µl
dH$_2$O fill to	100 µl

Table 39: Oligonucleotide sequences used for ligation-mediated PCR. *All three sense oligos pair with the GenRE-Adapt_AS oligo. S = sense (sequence 5' to 3'), AS = antisense (sequence 3' to 5').

Oligonucleotide	Sequence
BspHI-Adapt_S	CATGCACGTCGAATGCTACATGACACCAGGC
AseI-Adapt_S	TACACGTCGAATGCTACATGACACCAGGC
HindIII-Adapt_S	AGCTCACGTCGAATGCTACATGACACCAGGC
GenRE-Adapt_AS*	GTGCAGCTTACGATGTACTGTGGTCCG
GLuc-PCR_F	GAAGTCTTCGTTGTTCTCGGTGGGC

2.10.2 Preparation of ligation-mediated PCR

The ligation products were diluted in 90 µl H_2O and then used as the DNA template in a PCR reaction using the forward primer GLuc-PCR_F found in the GLuc sequence and using the antisense adapter oligonucleotide GenRE-Adapt_AS as the reverse primer. In order to capture possible PCR products with a length of several kilobases, a long-range temperature gradient program was implemented (Table 40). The standard PCR master mix was prepared either with or without DMSO, and containing 1 µl ligated DNA template (Table 2). The PCR products were separated on an agarose gel and visualized with EtBr. The sizes of expected PCR products can be calculated using the genomic sequence surrounding the AAVS1 safe harbor locus. Successful digestion and ligation of adapters was expected to yield a PCR product of at least 3.5 kbp from the HindIII digestion/ligation, at least 7 kbp from the BspHI digestion/ligation, and at least 9 kbp from the AseI digestion/ligation. When visible on the gel, successfully amplified products would be excised and purified and sent to Eurofins for Sanger sequencing to identify the locus at which the donor DNA was inserted into the genome.

Table 40: Long-range temperature gradient thermocycling program.

Step	Temperature (°C)	Duration (sec)	Number of cycles
Denaturation	98	30	1
Denaturation	98	20	
Annealing	55-70	30	45
Extension	72	150	
Extension	72	600	1

2.11 Double-control quantitative copy number PCR

The assessment of the insert copy number in each clone was carried out by qPCR. The primers used in the optimization of this protocol and the final method are found in Table 41.

62

Primer name	Sequence	Product size (bp)
Gaussia luciferase		
GLuc_Fprobe3	CCCCTTGATCTTGTCCACCT	299
GLuc_Rprobe4	GCACGCCCAAGATGAAGAAG	
Autosomal control gene		
hCHOP-F	CAGAACCAGCAGAGGTCACA	210
hCHOP-R	AGCTGTGCCACTTTCCTTTC	
Chromosome X genes		
xRBBP7_F	AAATTTCACTGACAGGGCCG	264
xRBBP7_R	GGCCATCTCAATTTGTCCCG	
xGATA1_F	CTGTTCTGGTAGCCTGTGGA	243
xGATA1_R	ACAGTTGAGGCAGGGTAGA	
xHPRT1_F	GGGCTAGACTTTTGAGGGACA	250
xHPRT1_R	AGTCCTAATCGGCCATTACTGA	
xTMSB15B_F	GTTGCTTTCAGTCTCTGCCC	244
xTMSB15B_R	GGGTAGCAGCAAACTCACAG	

The control plasmids were created using the hPOMC-GLuc donor plasmid as a backbone. The sequence encoding Gaussia luciferase was already incorporated in this plasmid. The sequence encoding a 200 bp fragment of the autosomal gene *CHOP* was amplified by PCR using the Phusion polymerase (Table 2), which produces PCR products with blunt ends. After thermal cycling, the products were separated on an agarose gel, visualized with EtBr, excised from the gel, and purified as described above. Then, the purified product was phosphorylated at 37 °C for 30 min with 1 µl T4 PNK reagent in 1x T4 ligation buffer. Next, the plasmid vector was digested with a single-cutting blunt-end enzyme (PmlI) and dephosphorylated (Table 8). Finally, the insert and vector were ligated, transformed, and amplified as described in the section 2.1.1. This plasmid now contained sequences encoding GLuc and a portion of *CHOP*, and was once again blunt-end cloned (restriction enzyme PsiI) with each of the ChrX genes as just described. Therefore, four plasmids were prepared for the optimization of this method in order to contain the sequences for Gaussia luciferase, and partial sequences of the autosomal gene *CHOP* and one each of *RBBP7*, *GATA1*, *HPRT1* or *TMSB15B* ChrX genes. Hereafter, these plasmids are referred to by their corresponding ChrX gene name (e.g. the "RBBP7 plasmid" contains the coding sequence for GLuc, a partial sequence of *CHOP*, and a partial sequence of *RBBP7*).

The "RBBP7 plasmid" was used for the final copy number analyses found in the results (section 3.4.2) and an example calculation for insert copy number is given below based on this plasmid. The optimization including the remaining plasmids was calculated in an analogous fashion. The qPCR plate prepared according to this method contained the control plasmid, the samples under investigation, and non-template controls for each sequence under investigation: GLuc, *CHOP* and *RBBP7*. Furthermore, control HapMap gDNA samples derived from a female and a male were included in each plate and analyzed with *CHOP* and *RBBP7* primer sets. These samples were not included in the GLuc measurement because luciferase is not encoded in the human genome. All samples were run in triplicate using the master mix described in Table 42 with the standard qPCR thermocycling program (Table 18).

Table 42: qPCR master mix composition for copy number analysis.

Reagent	Final concentration
2× Maxima SYBR Green Master Mix	1×
Primer F (10µM)	250 nM
Primer R (10µM)	250 nM
DNA (10 ng/µl)	10 ng

Once Ct values for each sample and sequence were collected, the sample amplification was normalized to the plasmid amplification within each gene analyzed:

$$(1) = normalization\ within\ gene = \left. 2^{Ct_{plasmid}-Ct_{sample}} \middle/ average\ (2^{Ct_{plasmid}-Ct_{plasmid}}) \right.$$

$$(2) = normalization\ relative\ to\ CHOP = \left. (1)_{sample(GLuc\ or\ RBBP7)} \middle/ (1)_{sample\ (CHOP)} \right.$$

$$(3) = normalization\ relative\ to\ RBBP7 = \left. (2)_{GLuc} \middle/ (2)_{RBBP7} \right.$$

For these calculations, the primer efficiency was normalized by comparing each gene of interest with the same plasmid. Furthermore, the normalization relative to *CHOP* designated which Ct value, within each sample, equals 2, since autosomal genes possess two alleles. The secondary normalization to *RBBP7* verified that the *CHOP* normalization results in the ChrX gene equals 2 in females and 1 in males. Female and male gDNA were always included as a positive control during qPCR analyses.

Section 2.11

2.12 Gene expression analysis

2.12.1 RNA extraction and reverse transcription

For gene expression analysis, RNA extraction was carried out with the ReliaPrep RNA Tissue Miniprep System (Promega) according to manufacturer's instructions. This kit is optimized for extractions of low-concentration RNA, which typical of motor neuron samples. RNA concentrations were measured with a NanoDrop. For "low concentration" samples with RNA concentrations less than 100 ng/µl, all available sample was included in the RT-PCR; for "high concentration" samples with concentrations greater than 100 ng/µl, a total of 2 µg RNA was reverse transcribed. Mix A was prepared for each RNA sample (Table 43) and run according to part A of the thermocycling protocol (Table 44). Next, mix B (Table 43) was added to each sample and the RT-PCR was carried out according to part B of the thermocycling program (Table 44).

Table 43: Reagent components for reverse transcription of RNA to cDNA.

Reagent	Final concentration	Volume (µl)
Mix A		
Oligo (dt) 18 (500ng/µl)	500 ng	1
RNA	up to 2 µg	x
Nuclease-free H$_2$O fill to		13
Mix B		
5× Buffer RT	1×	4
dNTPs (10 mM)	1 mM	2
Reverse Transcriptase (200U/µl)	200 U	1
or		
Revert Aid M-MUL (H-Minus) (200U/µl)	200 U	1

Table 44: Parts A and B of RT-PCR thermocycling program.

Step – Part A	Temperature (°C)	Duration (sec)
Poly(A) annealing	70	300
Reaction termination	4	300
Step – Part B	Temperature (°C)	Duration (sec)
Primer extension	25	600
cDNA synthesis	42	150 minutes
Reaction termination	70	600

2.12.2 cDNA preparation and qPCR

To prepare the cDNA samples for qPCR, the "high RNA concentration samples" (see above) were diluted to 4.4 ng/µl each. Conversely, the "low RNA concentration samples" (see above) were diluted 4 parts cDNA to 45 parts H_2O. The GLuc sequence and genes whose expressions are associated with iPSCs and MNs were amplified by qPCR (primers and corresponding product sizes in Table 45). The qPCR master mix for the gene expression measurement is described in Table 46 and the two thermocycling programs used are described in Table 47 and Table 18. The housekeeping genes *PPIA* and *RPS23* and the MN expression markers were amplified using the thermocycling program found in Table 47. The cDNA including GLuc and the genes associated with pluripotency was amplified using the thermocycling program found in Table 18. Samples were run in triplicate and analyzed with the $2^{-\Delta\Delta Ct}$ (delta delta cycle threshold) method using the geometric mean[101] of expression of the reference genes *PPIA* and *RPS23*. The formula to calculate the relative fold gene expression against the reference sample is as follows, whereby goi = gene of interest, and ref gene = reference gene:

$$\frac{2^{Ct\ goi_{control}-Ct\ goi_{sample}}}{2^{Ct\ ref\ gene_{control}-Ct\ ref\ gene_{sample}}}$$

Statistical differences in gene expression were analyzed using the unpaired 2-tailed t-test at 95% confidence level, in GraphPad Prism.

Table 45: Primers used for gene expression analysis of Gaussia luciferase, and genes associated with pluripotency and motor neurons. Sequences and product sizes are specified for each amplicon.

Primer name	Sequence	Product size (bp)
Reference genes		
PPIA_F	GCCAAGACTGAGTGGTTGGAT	75
PPIA_R	GGCCTCCACAATATTCATGCC	
RPS23_F	ACAGGATGGGCAAGTGTCGT	70
RPS23_R	CACTTCTGGTCTCGTCGGTG	
Gaussia luciferase expression		
GLuc_Fprobe3	CCCCTTGATCTTGTCCACCT	299
GLuc_Rprobe4	GCACGCCCAAGATGAAGAAG	
Pluripotency expression		
qoct4_F	CGAGAAGGATGTGGTCCGAG	213
qoct4_R	GGGAAAGGGACCGAGGAGTA	
qNANOG_F	TGAGTGTGGATCCAGCTTGT	284
qNANOG_R	TTTCTTGACCGGGACCTTGT	

qLIN28A_F	TGTAAGTGGTTCAACGTGCG	283
qLIN28A_R	TGTCTCCTTTTGATCTGCGC	
hSOX2_F	TGGACAGTTACGCGCACAT	100
hSOX2_R	CGAGTAGGACATGCTGTAGGT	
Motor neuron marker expression		
OLIG2_F	TATAGATCGACGCGACACCAG	93
OLIG2_R	GGACCCGAAAATCTGGATGC	
HB9_F	GAGACCCAGGTGAAGATTTGGT	70
HB9_R	GCTCTTTGGCCTTTTTGCTGC	
ISLET1_F	GGATTTGGAATGGCATGCGG	135
ISLET1_R	CATTTGATCCCGTACAACCTGA	
CHAT_F	AAGGAGTAGGAGCCGAGCAT	79
CHAT_R	CACCCGAATTTCCAGAGGTCG	

Table 46: Master mix for gene expression analysis with SYBR green.

Reagent	Final concentration
2× Maxima SYBR Green Master Mix	1×
Primer F (10µM)	250 nM
Primer R (10µM)	250 nM
DNA (4.4 ng/µl)	20 ng

Table 47: Thermocycling protocol for gene expression in MNs

Step	Temperature (°C)	Duration (sec)	Number of cycles
Denaturation	95	600	1
Denaturation	95	10	40
Annealing	67.5	30	
Melting curve	95	10	1
	65	5	1
	95	Ramp up from 65 °C	

2.13 Differential centrifugation

Cultured cells were prepared for differential centrifugation according to a protocol adapted from Yu *et al*, 2013[102]. The instructions to prepare the subcellular fractionation buffer (SF buffer) and the nuclear lysis buffer (NL buffer) are found in Table 48 and Table 49, respectively. An aliquot of cell culture medium in which the cells were cultivated was removed and the total volume of medium was recorded for later luciferase activity quantification. The temperature-sensitive and time-sensitive buffer components were added shortly before starting the centrifugation process. Cells were washed twice with ice-cold PBS and while on ice, 500 µl of SF buffer was immediately added into each 100 mm plate. If multiple samples were

collected, one sample was processed at a time. Then, the homogenates were transferred to a 1.5 ml tube and were agitated at 4 °C for 30 min at 30-50 rpm on a tube roller and then centrifuged at 720 x g at 4 °C for 5 min. The supernatant (S1) was carefully transferred to a new 1.5 ml tube. Next, the pellet (P1) was washed with 500 µl of SF buffer and resuspended with a pipette. An aliquot was removed to measure activity (P1). Then, the sample was centrifuged at 720 x g at 4 °C for 10 min.

The supernatant (S2) was removed and the pellet was resuspended in 500 µl NL buffer. The sample was transferred to a homogenization glass was homogenized using the Potter homogenizer by moving the glass up and down for 20 strokes at 3000 rpm. The homogenate was incubated at 4 °C for 15 min. An aliquot of this sample was removed to measure luciferase activity (P2). Then, the sample was centrifuged at 10,000 x g at 4 °C for 10 min. The supernatant (S3) was carefully transferred to a new 1.5 ml tube and an aliquot was removed to measure luciferase activity. S3 corresponds to the cytosolic and membrane fraction. Next, the pellet (P3) was resuspended in 500 µl SF buffer to measure luciferase activity. The supernatant (S3) was centrifuged in an ultracentrifuge at 100,000 x g at 4 °C for 1 h. The supernatant (S4) was carefully transferred to a new 1.5 ml tube. S4 corresponds to the cytosolic fraction. Then, the pellet was washed with 500 µl of SF buffer and resuspend by pipetting, an aliquot was removed to measure luciferase activity (P4). Finally, the resuspended pellet was ultracentrifuged at 100,000 x g at 4 °C for 1 h. The supernatant (S5) was removed and the pellet was resuspended in NL buffer (P5).

The luciferase activity and protein concentration were measured in P1 - P5 and S1 - S5. The luciferase activity was measured as described in section 2.5 and corrected for specific activity by normalization to the sample's protein concentration. The protein concentration was determined by the Bradford method[103,104] and absorption values were measured at 595 nm with the FLUOstar Omega microplate reader. Each aliquot was diluted 1:7 with dH$_2$O. Next, 10 µl of diluted sample was mixed with 90 µl Bradford reagent and the concentration of each sample was calculated against a standard BSA dilution series. The relevant measurements of this assay are the specific luciferase activities in the cellular debris (aliquot P1), the cytosolic fraction (aliquot S3), and the membrane fraction (aliquot P5).

Table 48: Composition of subcellular fractionation buffer for subcellular fractionation.

Reagent	Final concentration	1× solution
Sucrose	250 mM	4.28 g

Section 2.13

HEPES (pH 7.4, [1 M])	20 mM	1 ml
KCl	10 mM	0.0373 g
MgCl₂ [1 M]	1.5 mM	75 µl
EDTA [0.5 M]	1 mM	100 µl
EGTA [0.5 M]	1 mM	100 µl
H₂O fill to		50 ml
At time of use add the following into 10 ml of SF buffer:		
DTT [1 M]	1 mM	10 µl
PI cocktail [40×]	1×	250 µl

Table 49: Composition of nuclear lysis buffer for subcellular fractionation.

Reagent	Final concentration	1x solution
Tris HCl (pH 8, [1 M])	50 mM	2.5 ml
NaCl [1 M]	150 mM	7.5 ml
NP-40 [20 %]	1 %	2.5 ml
Sodium deoxycholate [10 %]	0.5 %	2.5 ml
SDS [10 %]	0.1 %	0.5 ml
At time of use add the following into 10 ml of NL buffer:		
PI cocktail [40×]	1x	250 µl
Glycerol [10%]	1 %	1 ml

2.14 Immunofluorescence

2.14.1 Cell treatments

Standard staining protocol

Cells were plated on cover slips in 24-well plates with a surface coating of Matrigel (iPSCs or iPSC-derived cells) or 1× PLL (SIMA) and cultivated under standard conditions at 37 °C and 5 % CO_2 (see section 2.3.1) After the cells had adhered to the plate, they were fixed with 4 % paraformaldehyde for 10 min at RT. All subsequent steps, except for the primary antibody incubation, were performed on a rocking device. The coverslips were washed 3x for 10 min in PBS. The coverslips were either stored in PBS at 4 °C or the IF protocol was continued immediately. The cells were permeabilized in 0.25 % PBS-T for 10 min at RT. The permeabilization buffer was aspirated and the cells were blocked in blocking solution for 60 min at RT. The primary antibodies were diluted in blocking solution according to Table 50. In order to reduce antibody usage, the coverslips were removed from the 24-well plate and incubated, cell-side down, in 25 µl antibody dilution ON at 4 °C. The next day, the cover slips were transferred back into 24-well plates, cell-side up, and excess antibodies were removed by washing 3x for 10 min in PBS. The secondary antibodies were diluted in blocking solution

according to Table 50 and incubated at RT for 1 h. This step and all subsequent steps were carried out in the dark to protect the fluorescent tags. The secondary antibody solution was aspirated and the cells were incubated in 300 nM DAPI, diluted in PBS, for 5 min. The cover slips were then washed 3x for 10 min in PBS. The cover slips were mounted on glass microscope slides using 3.2 µl Vectashield Fluorescence Mounting Medium and sealed by applying clear nail polish to the perimeter of the cover slip.

Table 50: Primary and secondary antibodies used in immunofluorescence.

Antibody target (host)	Dilution
Primary antibodies	
Gaussia luciferase (NEB) (rabbit)	1:1000
Gaussia luciferase (SC) (rabbit)	1:50
Golgi97 (mouse)	1:50
Leptin HPA (Golgi apparatus)	5 µg/ml
GM130 (mouse)	1:250
Chromogranin A (CgA) (goat)	1:50
Secretogranin II (SgII) (rabbit)	1:1000
Chromogranin A (CgA) (mouse)	1:100
Secretogranin II (SgII) (mouse)	1:500
Synaptophysin (Syp) (mouse)	1:200
GAPDH (mouse)	1:50
Islet1 (mouse ab86501)	1:500
Islet1 (mouse ab86472)	1:200
Secondary Alexa-fluor labelled antibodies	
Anti-mouse 488	1:1500
Anti-rabbit 488	1:750
Anti-rabbit 568	1:1000

Table 51: Buffers used in immunofluorescence.

4 % Paraformaldehyde	
PFA	40 g
PBS fill to	1000 ml
Adjust to pH 6.9	
Permeabilization buffer	
Triton-X	0.25 %
in PBS	
Immunofluorescence Blocking solution	
FCS	5 %
BSA	1 %

Section 2.14

Triton-X	0.25 %
in PBS	

Alternative Golgin-97 staining protocol

Cells were fixed in 2% PFA and then washed twice in 120 mM NaxHxPO4, pH 7.4 and twice in high-salt PBS (0.1% Triton X-100, 150 mM NaCl and 3.3 mM NaxHxPO4, pH 7.4 in PBS). Cells were blocked in goat serum dilution buffer (GSDB; 10% goat serum, 150 mM NaCl, 6.6 mM NaxHxPO4 and 0.1% Triton X-100 in PBS) for 20 min and then primary antibodies diluted in GSDB were incubated with cells for 1 h at RT. Cover slips were washed three times in high-salt PBS for 10 min and secondary antibodies diluted in GSDB were incubated with cells for 1 h at room temperature. Prior to mounting, cells were washed twice in high-salt PBS for 5 min and twice in 120 mM NaxHxPO4 for 5 min (protocol communicated by Dr. Michael Krauss, Leibniz-Forschungsinstitut für Molekulare Pharmakologie).

Lectin HPA protocol

Cells were fixed in 4 % PFA and washed three times in PBS. Cells were permeabilized in 0.25 % PBS-T for 5 min at RT and then incubated for 15 min in 5µg/mL of Lectin HPA Alexa Fluor R 647 conjugate to visualize the Golgi apparatus.

2.14.2 Image acquisition

Images of cells were taken with a Leica DM6 B with a CTR6 LED, using an HC Plan APO 20×/0.70 objective and an HC PL APO 63×/1.40-0.60 oil objective. Confocal microscopy was carried out on a Zeiss LSM 780 T-PMT Observer.Z1 inverted modular microscope using the Plan-Apochromat 63x / 1.40 Oil DIC M27 objective. The lasers used were the 405 nm Diode laser, and the 488 nm and 543 nm Argon lasers. Z-stack scans through the cell were taken through the ZEN 2012 SP5 program.

2.14.3 Colocalization analysis

Cell images were compiled and analyzed using Fiji (Fiji Is Just ImageJ)[105]. The GLuc localization confirmation can be carried out with immunofluorescence and the colocalization analysis carried out using high resolution stacked images taken with a confocal microscope. The colocalization was analyzed by selecting small regions within the image where the naturally occurring protein can be identified and then measuring the fluorescence levels through the image stack in this section of both the naturally occurring protein and GLuc (selected

regions are marked with arrows on each image overlay (sections 3.4.3 and 3.9.3)). The graphical analysis was done with the Plot Z-axis Profile function of Fiji, whereby the two fluorescence intensities through a section of the cell were plotted against each other as a function of distance along the Z-axis. The Pearson's correlation coefficient, corrected for mean intensity values of both fluorescence channels by the overlap coefficient, k1, and k2[106] (termed overlap coefficient), was calculated in the same section of the cell that was previously analyzed in the Z-axis plot using the Fiji plugin JACoP[93]. Statistical significance of the difference between two correlation coefficients was calculated using the unpaired 2-tailed t-test at 95% confidence level, in GraphPad Prism.

2.15 Luciferase release assay

In preparation for the luciferase release experiment the appropriate wells in the 96-well plates were coated with either $1\times$ PLL for SIMA cells or Matrigel for iPSCs and iPSC derived cells. In the MoN-Light BoNT assay, cells were incubated with 100 pM BoNT/A for 48 h at 37 °C before incubating the cells in control/stimulation buffers (Na$^+$-HBS and K$^+$-HBS buffers, respectively Table 52). The luciferase release assay was carried out with the following steps: the medium was aspirated off the cells which were washed once with fresh medium. The fresh medium was aspirated and the cells were incubated for 10 min at 37 °C in fresh medium. The medium was aspirated and 100 µl of appropriate control or depolarization buffer was added to cells. hPOMC-GLuc SIMA cells were incubated for 3 min at 37 °C with the test buffers; iPSC-derived hPOMC-GLuc and no tag GLuc motor neurons were also incubated for 3 min at 37 °C; iPSC-derived VAMP2-GLuc motor neurons were incubated for 2 h at 24 °C. The buffer supernatants were transferred to 1.5 ml tubes, which were centrifuged for 5 min at 2000 rpm to remove any cell debris and the clean supernatant was transferred to new tubes. Next, 20 µl of the clean supernatant was transferred to a white-walled white-bottomed 96-well plate for luciferase activity detection. The luminescence in the supernatant of the cells exposed to each buffer was measured over 9.9 s with automatic injection of 100 µl coelenterazine as described in section 2.5. The background luminescence of the GLuc-expressing cells that remained on the white-walled clear-bottomed 96-well growth plates was also measured with the automatic injection of 100 µl coelenterazine, which was used as the "lysate" measurement.

Table 52: Components of the standard Na$^+$-HBS (control) and K$^+$-HBS (depolarization) buffers. [mM] = concentration in millimolar

Na$^+$-HBS buffer [mM]	Component	K$^+$-HBS buffer [mM]

20	**HEPES, pH 7.4**	20
136	**NaCl**	40.7
4.7	**KCl**	100
1.25	**CaCl2**	1.25
1.25	**MgSO$_4$**	1.25

For each supernatant and lysate sample, the luminescence data was calculated using the mean of 89 luminescence unit (LU) readings from 1 - 9.9 sec. For each well, mean supernatant LU was divided by mean lysate LU to normalize any potential variations in luciferase expression. For the IMR90-4 VAMP2-GLuc analyses, this normalized value was multiplied by 100 to determine the luciferase release as the percentage of lysates. For the IMR90-4 hPOMC-GLuc analyses, each sample's LU value was expressed as the percentage difference of the average normalized Na$^+$-HBS and K$^+$-HBS values for all samples. Finally the average and standard deviation of triplicate samples was calculated. Statistical significance of differences in luciferase activity among the conditions tested were determined by t-test using the Holm-Sidak method in GraphPad Prism. Statistical significance of difference in luciferase activity among individual clones, regardless of buffer treatment were analyzed by two-way ANOVA, and Sidak's post-hoc multiple comparison test was applied.

3 Results

3.1 Production of CRISPR/Cas9 nuclease and donor plasmids

The specific cloning of plasmids was carried out according to the protocol in the Materials and Methods section 2.1. The CRISPR/Cas9 gene editing was applied in SIMA neuroblastoma cells and IMR90-4 iPSCs through co-transfection of two unique plasmids. The first is the eSpCas9(1.1)_AAVS1-T2 plasmid, a ligated product of the eSpCas9(1.1) plasmid and annealed oligonucleotides coding for the AAVS1-T2 gRNA. The second is the pAAVS1-P-MCS donor plasmid, consisting of the homology arms matching the sequence flanking the AAVS1 safe harbor locus, and between these two arms the DNA sequence to be integrated into the genome. All cloned sequences can be found in the Appendix 6.3.

3.1.1 Integration of eSpCas9(1.1) plasmid with AAVS1-T2 gRNA

The sequence encoding the AAVS1-T2 gRNA was produced by annealing two inverse complementary oligonucleotides that contained non-complementary sequences to produce 5' and 3' overhangs pertaining to BbsI digested double-stranded eSpCas9(1.1) plasmid DNA. Next, the annealed AAVS1-T2 oligonucleotides were ligated into the linear BbsI digested eSpCas9(1.1) plasmid. All picked clones were shown by PCR to have incorporated the sgRNA segment, plasmids 1-3 were sent for Sanger sequencing (Figure 23). Sequences 1 and 3 were 100% identical to the expected sequence and in the correct orientation. The sequence analysis of plasmid 2 contained too many ambiguities to confirm the correct insertion (see Appendix, 6.4.1). Plasmid eSpCas9(1.1)_AAVS1-T2 #1 was therefore maxi-prepped and used for all subsequent experiments.
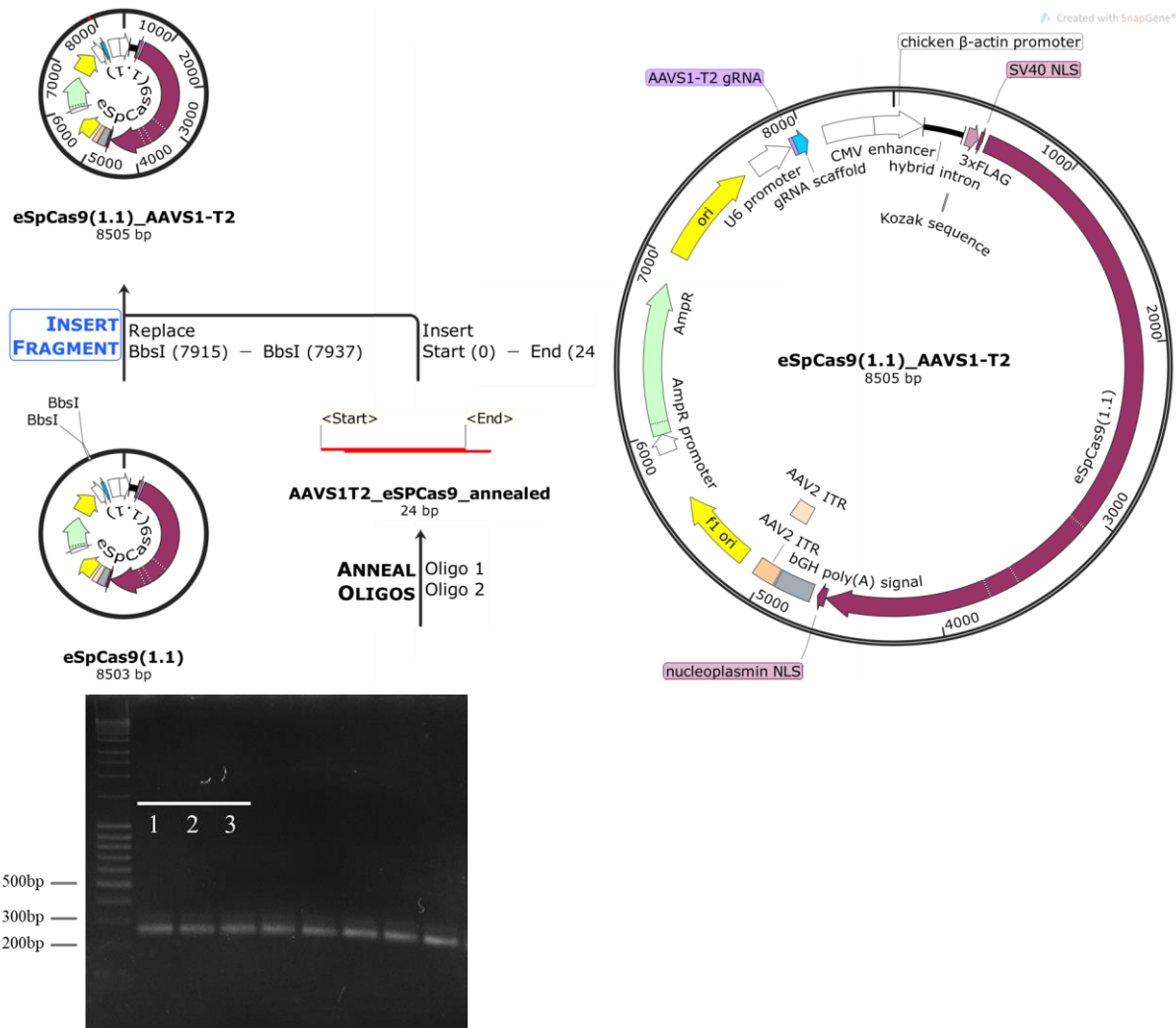
Figure 23: Integration of eSpCas9(1.1) plasmid with AAVS1-T2 gRNA. Upper panel: History of the cloning method from eSpCas9(1.1) to eSpCas9(1.1)_AAVS1-T2 and plasmid map of the Cas9 plasmid with the integrated AAVS1-T2 gRNA sequence. Lower panel: Gel confirmation of sgRNA (gRNA) sequence integration in the eSpCas9(1.1) plasmid. Expected product size 255 bp; Primers: Forward eSpCas9_gRNA_F-seq; Reverse AAVS1_T2_eSpCas9_AS; sequencing: eSpCas9_gRNA_F-seq.

3.1.2 Production of pAAVS1-P-MCS donor plasmids

Two plasmids from the pAAVS1-P-MCS donor backbone were prepared in parallel with two different promoters, the Ef1α promoter and the Ef1-HTLV promoter.

3.1.2.1 pAAVS1-P-MCS with human Ef1α promoter

The human Ef1α promoter has been shown to be a strong and consistent promoter active in many cell types, even post-differentiation[107–109]. The promoter sequence was amplified from gDNA of human origin using flanking primers with 5' overhangs corresponding to the SpeI and PacI restriction enzyme recognition sites. The promoter product and the pAAVS1-P-MCS backbone were digested with the PacI and SalI restriction enzymes and then ligated together. The insertion of the promoter into the plasmid was verified by digestion of the plasmid with the same restriction enzymes. The expected product sizes were 5.5 kbp and 1.5 kbp. Plasmids 1, 3,

5, and 6 all appear to have the promoter ligated into the plasmid backbone (Figure 24). Only the plasmid #1 was sent for Sanger sequencing, with sequencing primers flanking either side of the Ef1α segment, to allow the entire sequence to be controlled, despite its size. The promoter was confirmed by Sanger sequencing to be inserted in the correct orientation and to be 100 % identical to the expected sequence. The plasmid was maxi-prepped and used as a backbone for the subsequent Ef1α-hPOMC-GLuc plasmid (see Appendix, 6.4.2).
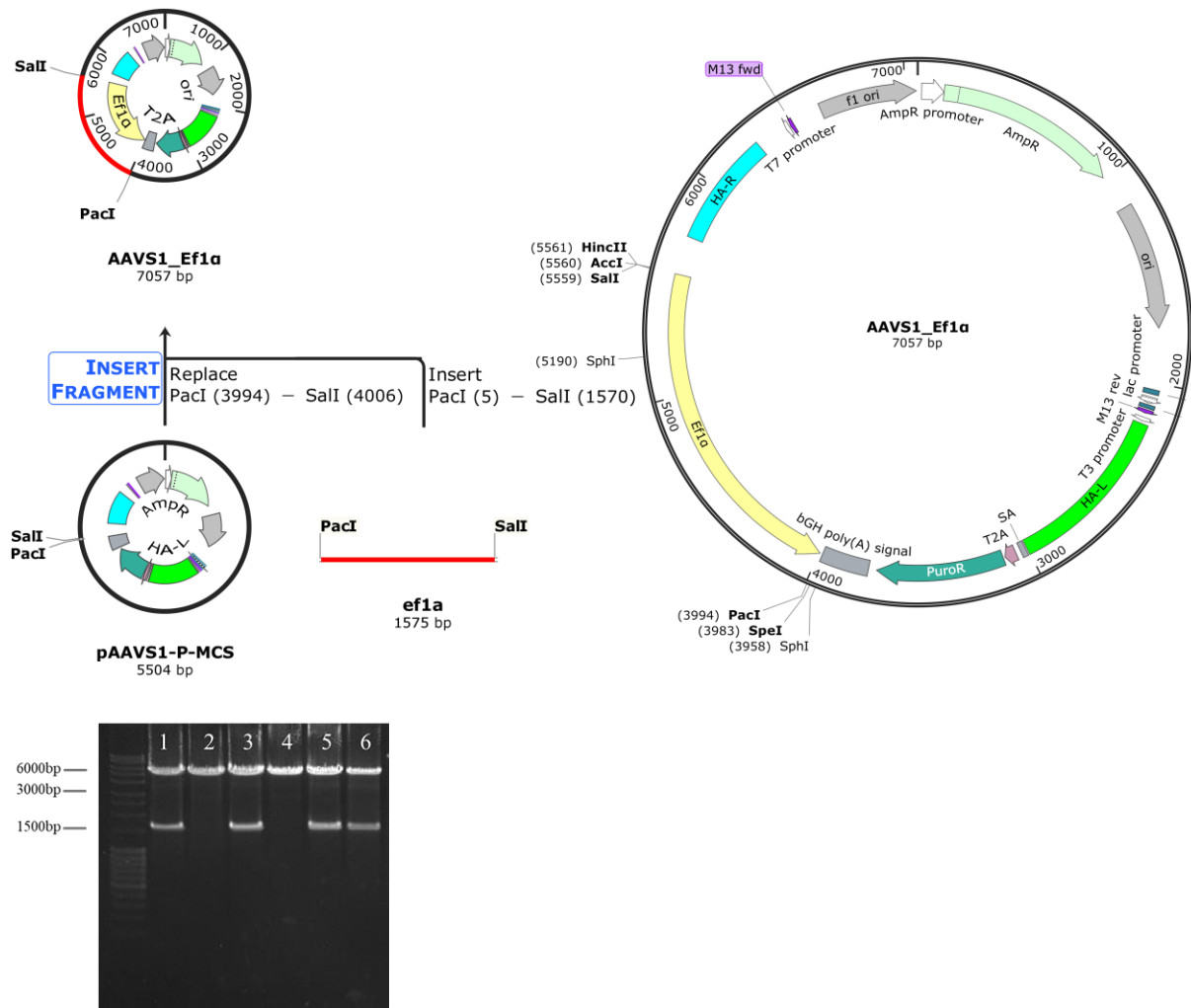
Figure 24: pAAVS1-P-MCS with human Ef1α promoter. Upper panel: History of the cloning method from pAAVS1-P-MCS to pAAVS1_Ef1α and plasmid map of the donor plasmid with the Ef1α promoter. Lower panel: Gel of digested mini-prepped clones. Expected product sizes of the digested plasmid are 5.5 kbp and 1.5 kbp.

3.1.2.2 pAAVS1-P-MCS with Ef1-HTLV promoter

The DNA sequence of the second promoter plasmid, Ef1-HTLV, was selectively amplified from the pNiFty3-SEAP plasmid (now called pUNO[97]) using flanking primers with 5' overhangs corresponding to the SpeI and PacI restriction enzyme recognition sites. The promoter sequence and the pAAVS1-P-MCS backbone were digested with the PacI and SalI

restriction enzymes and then ligated together (upper panel Figure 25). A selection of the resulting clones were digested with the restriction enzymes PacI and SalI. Only plasmids 5 and 6 contained an insert of the expected size of 5.5 kbp (Figure 25). Sanger sequencing of plasmid 5 (coded as 43EH18) revealed that the promoter was inserted in the correct orientation and the sequence was 100 % identical to the expected sequence, whereas plasmid 6 (coded as 43EH17) contained a mutation (Appendix, 6.4.3). Plasmid 5 was maxi-prepped and all subsequent tagged-GLuc plasmids were constructed with this backbone.
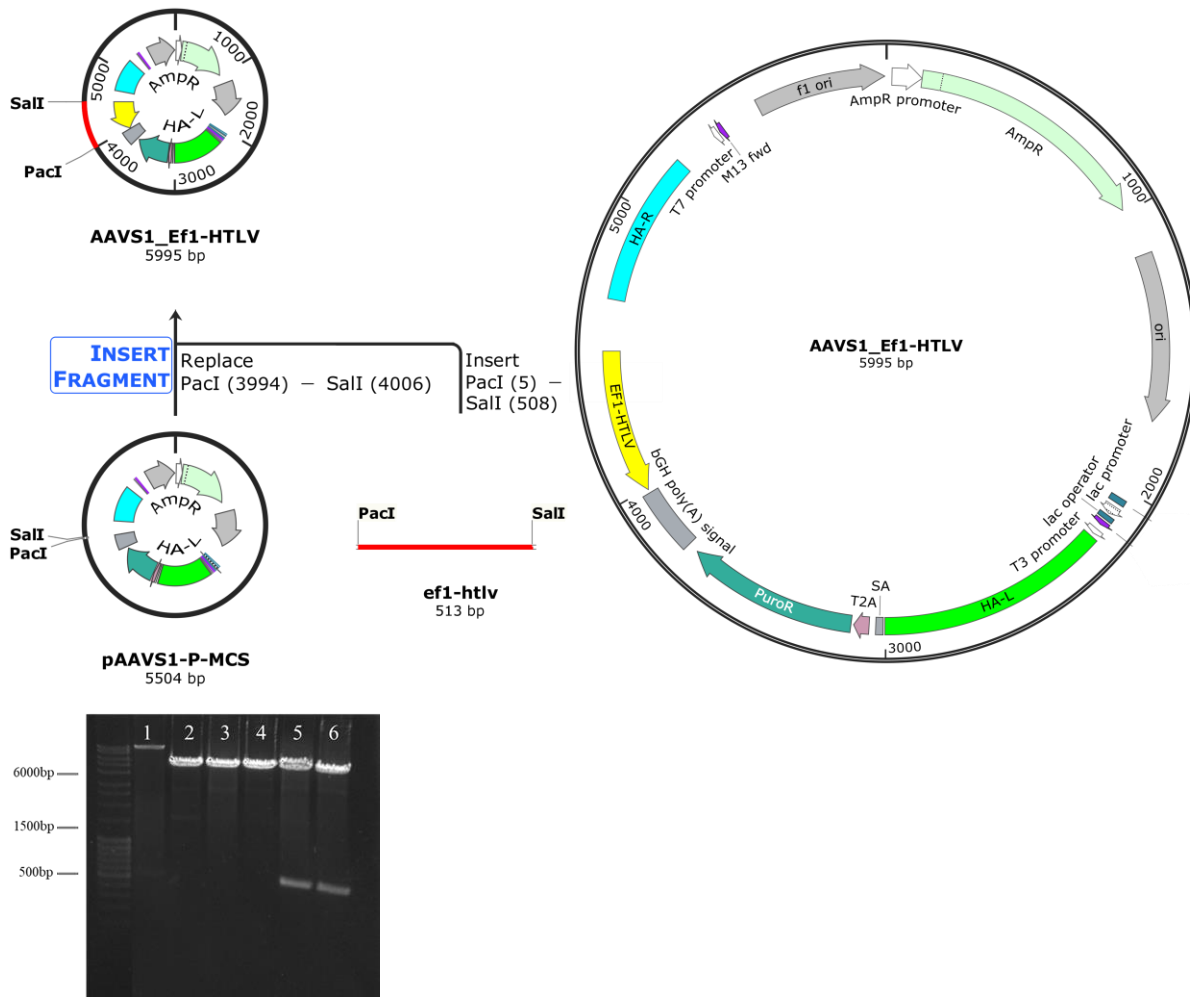
Figure 25: pAAVS1-P-MCS with Ef1-HTLV promoter Upper panel: History of the cloning method from pAAVS1-P-MCS to pAAVS1_Ef1-HTLV and plasmid map of the donor plasmid with the Ef1-HTLV promoter. pAAVS1_Ef1-HTLV was the backbone for all subsequent GLuc-expressing plasmids. Lower panel: Gel of digested mini-prepped clones. Expected product sizes of the digested plasmid are 5.5 kbp and 500 bp.

3.1.2.3 hPOMC-GLuc donor plasmids

The hPOMC-GLuc fusion sequence was amplified using the recently described pcDNA3-hPOMC1-26-GLuc plasmid[2] as a template and flanking primers with 5' overhangs corresponding to the SpeI and PacI restriction enzyme recognition sites. The resulting PCR product was digested with SpeI and PacI and was cloned in parallel into both the pAAVS1-P-

MCS Ef1α backbone (Figure 26) and the pAAVS1-P-MCS Ef1-HTLV backbone (Figure 27) cleaved beforehand with SpeI and PacI.

pAAVS1-P-MCS Ef1α hPOMC-GLuc

The plasmids extracted from the picked clones were digested with NcoI, a restriction enzyme recognition sequence that occurs twice in the hPOMC-GLuc construct, but not in the Ef1α hPOMC-GLuc backbone. The expected fragment sizes of 350 bp and 7.05 kbp were identified in clones 3, 4, and 6, which were sent for sequencing (lower panel Figure 26). Sanger sequencing revealed that the inserts were in the correct orientation and no mutations were incorporated during the PCR amplification (Appendix, 6.4.4). Plasmid 3 was maxi-prepped and used for all subsequent transfections.

Section 3.1
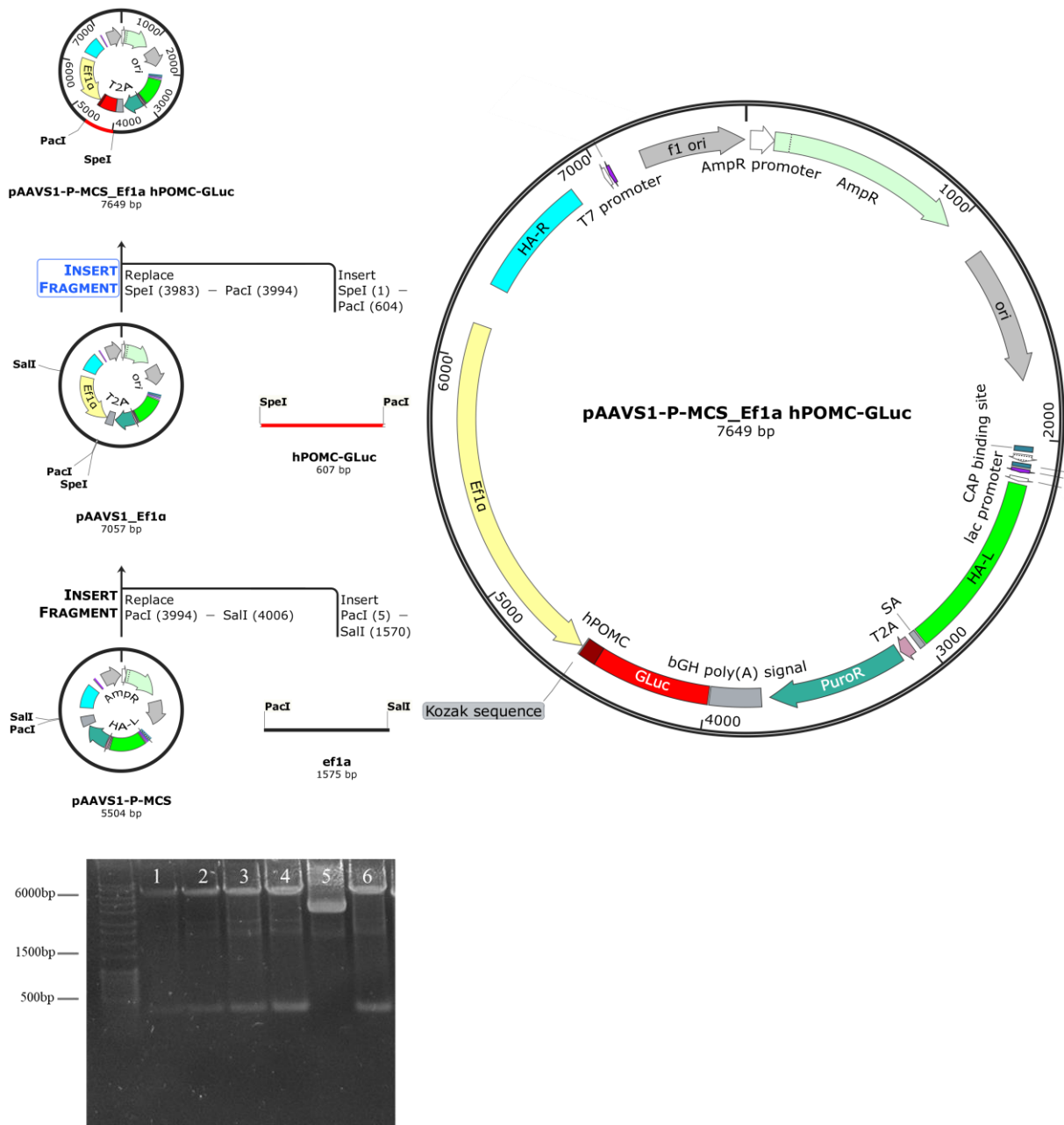
Figure 26: pAAVS1-P-MCS_Ef1a with hPOMC-GLuc Upper panel: History of the cloning method from pAAVS1-P-MCS to pAAVS1_Ef1α_hPOMC-GLuc and plasmid map of the pAAVS1_ Ef1α _hPOMC-GLuc donor plasmid. Lower panel: Gel of digested mini-prepped clones. Expected product sizes of the digested plasmid are 7 kbp and 350 bp.

pAAVS1-P-MCS Ef1-HTLV hPOMC-GLuc

The plasmids extracted from the picked clones were also digested with NcoI, a restriction enzyme recognition sequence that occurs twice in the hPOMC-GLuc construct, but not in the Ef1-HTLV hPOMC-GLuc backbone. The expected fragment sizes of 350 bp and 5.9 kbp were identified in all digested plasmids (lower panel Figure 27). In plasmids 1, 2 and 3, Sanger sequencing revealed that the inserts were in the correct orientation and no mutations were incorporated during the PCR amplification (Appendix, 6.4.5). Plasmid 2 was maxi-prepped and used for all subsequent transfections.
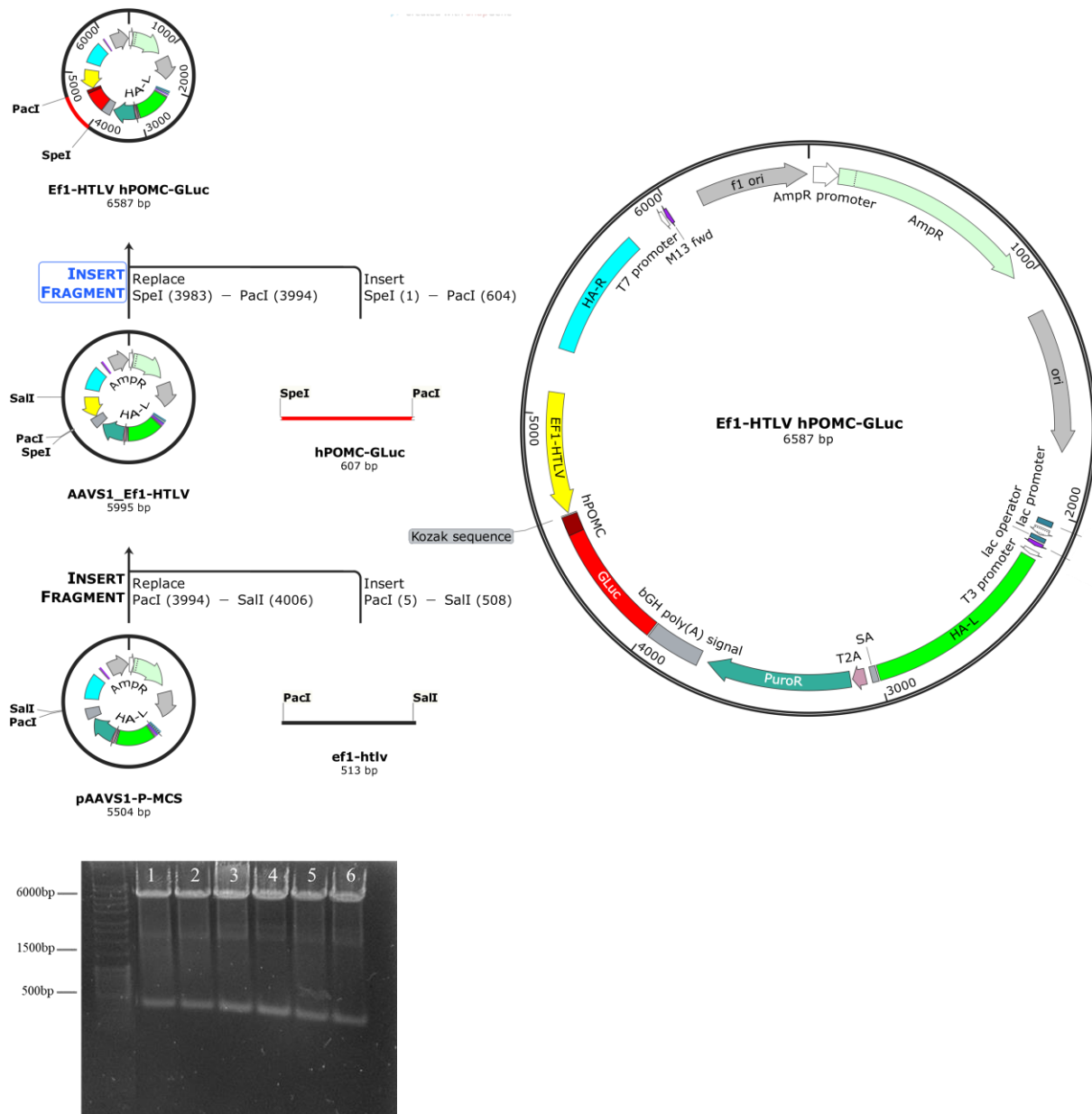
Figure 27: pAAVS1-P-MCS_Ef1-HTLV with hPOMC-GLuc Upper panel: History of the cloning method from pAAVS1-P-MCS to pAAVS1_Ef1-HTLV_hPOMC-GLuc and plasmid map of the pAAVS1_Ef1-HTLV_hPOMC-GLuc donor plasmid. Lower panel: Gel of digested mini-prepped clones. Expected product sizes of the digested plasmid are 5.9 kbp and 350 bp.

Transient transfection of Ef1α and Ef1-HTLV hPOMC-GLuc plasmids in SIMA cells

Upon transient transfection of SIMA neuroblastoma cells with the Ef1α and Ef1-HTLV hPOMC-GLuc donor plasmids, no luciferase activity was detected in cells transfected with the Ef1α promoter-based plasmid, whereas cells transfected with the Ef1-HTLV hPOMC-GLuc did express luciferase (Figure 32). It was assumed that the expression under control of the Ef1α promoter was inefficient, so this construct was no longer used and only the pAAVS1-P-MCS Ef1-HTLV promoter construct was used for downstream cloning and in the rest of this project.

Section 3.1

3.1.2.4 Additional donor plasmids with pAAVS1-P-MCS Ef1-HTLV backbone

no tag GLuc donor plasmid

The no tag GLuc sequence was amplified from the pcDNA3-hPOMC1-26-GLuc plasmid described above using flanking PCR primers with 5' overhangs corresponding to the SpeI and PacI restriction enzyme recognition sites, excluding the hPOMC sorting signal. The PCR product was digested with SpeI and PacI and cloned into the pAAVS1-P-MCS Ef1-HTLV donor plasmid backbone cleaved with the same restriction enzymes (upper panel Figure 28). The plasmids extracted from the picked clones were digested with SpeI and PacI restriction enzymes. The expected fragment sizes of 5.9 kbp and 515 bp were identified in all clones except clone 5 (Figure 28 lower panel). In clones 1, 3, and 7 Sanger sequencing revealed that the insert sequences were in the correct orientation and that no mutations were incorporated during the PCR amplification (Appendix, 6.4.6). Clone 3 was chosen for expansion and maxi-prep, and this plasmid was used for all relevant transfections.
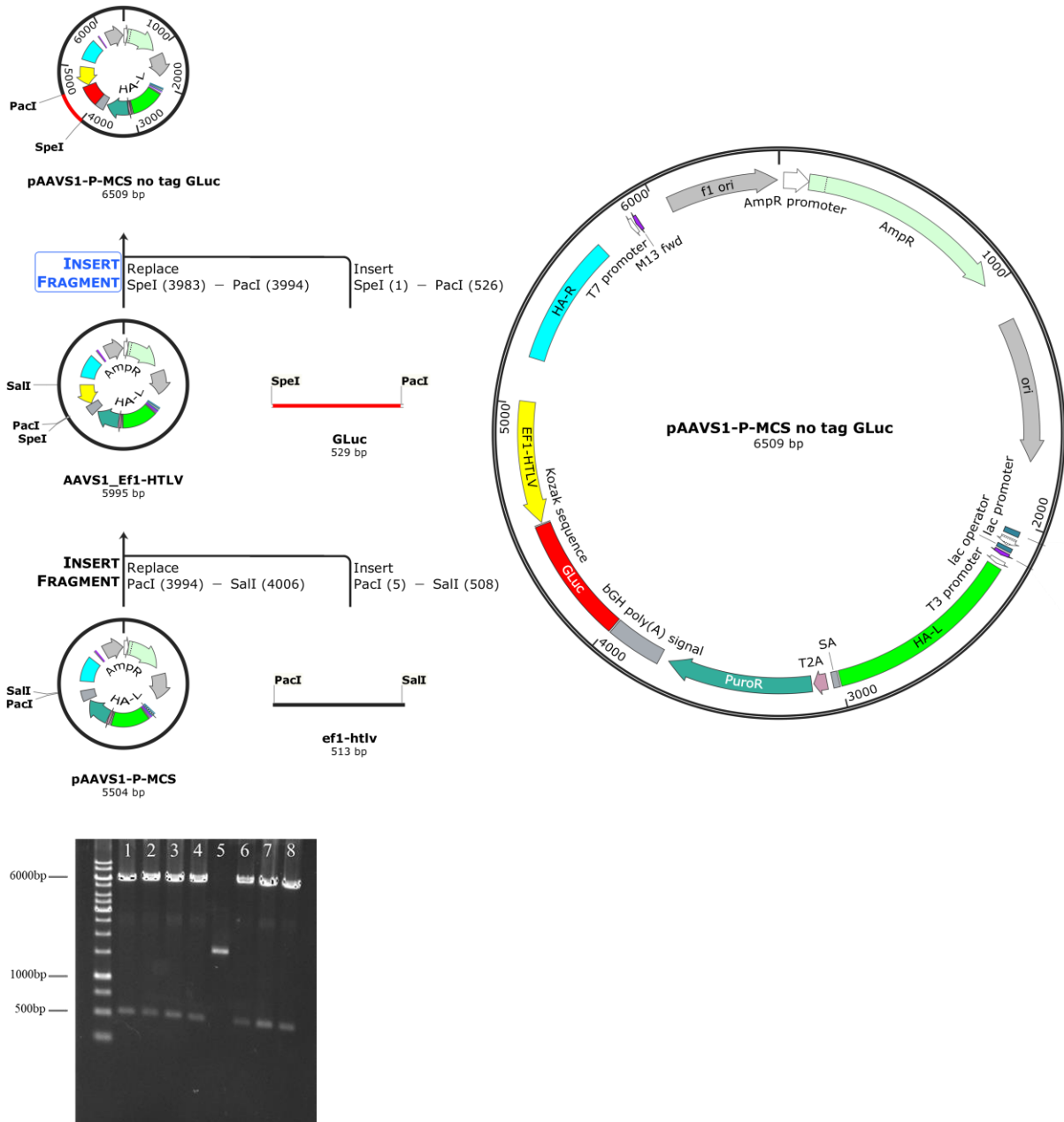
Figure 28: pAAVS1-P-MCS_Ef1-HTLV with no tag GLuc Upper panel: History of the cloning method from pAAVS1-P-MCS to pAAVS1_Ef1-HTLV no tag GLuc and plasmid map of the pAAVS1_Ef1-HTLV no tag GLuc donor plasmid. History and plasmid maps were created with SnapGene. Lower panel: Gel of digested mini-prepped clones. Expected product sizes of the digested plasmid with SpeI and PacI are 5.9 kbp and 515 bp.

CgA-GLuc donor plasmid

The CgA-GLuc fusion sequence was amplified using the recently described pcDNA3-hSCgA-GLuc plasmid[2] as a template and flanking PCR primers with 5' overhangs corresponding to the SpeI and PacI restriction enzyme recognition sites. The resulting product was digested with SpeI and PacI and was cloned into the pAAVS1-P-MCS Ef1-HTLV backbone cleaved with SpeI and PacI (upper panel Figure 29). The plasmids extracted from the picked clones were digested with SpeI and PacI restriction enzymes. The expected digestion

Section 3.1

products of 5.9 kbp and 1.9 kbp were identified in all plasmids (lower panel Figure 29). In clones 1, 2, and 3, Sanger sequencing revealed that the insert sequences were in the correct orientation and no mutations were incorporated during the PCR amplification (Appendix, 6.4.7). Plasmid 3 was expanded, maxi-prepped and used for all subsequent experiments.
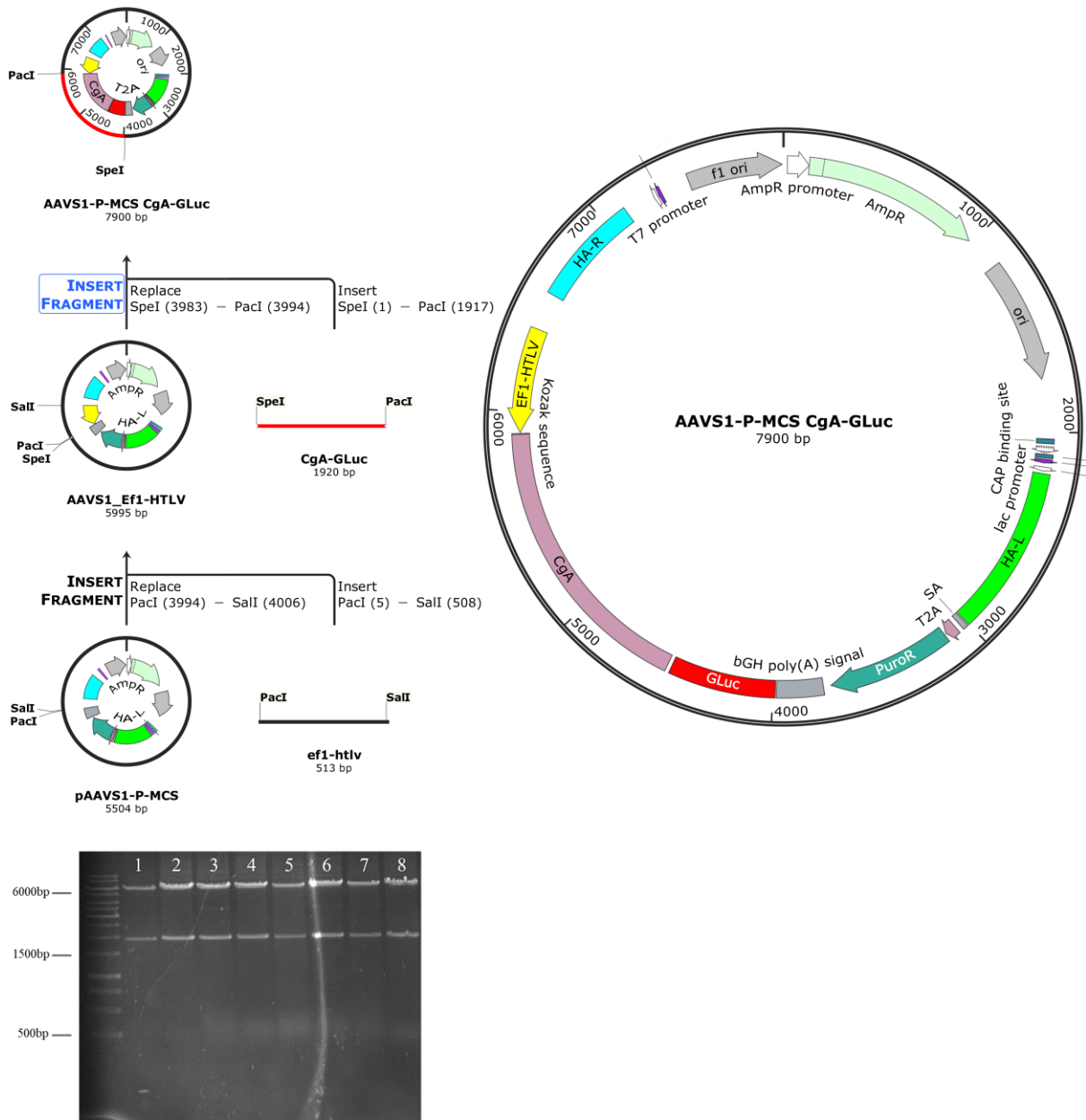
Figure 29: pAAVS1-P-MCS_Ef1-HTLV with CgA-GLuc Upper panel: History of the cloning method from pAAVS1-P-MCS to pAAVS1_Ef1-HTLV_CgA-GLuc and plasmid map of the pAAVS1_Ef1-HTLV_CgA-GLuc donor plasmid. Lower panel: Gel of digested mini-prepped clones. Expected product sizes of the digested plasmid with SpeI and PacI are 5.9 kbp and 1.9 kbp.

SgII-GLuc donor plasmid

The SgII-GLuc fusion sequence was amplified using the recently described pcDNA3-hSgII-GLuc plasmid[2] as a template, using flanking PCR primers with 5' overhangs corresponding to the SpeI and PacI restriction enzyme recognition sites. The resulting product

was digested with SpeI and PacI and was cloned into the pAAVS1-P-MCS Ef1-HTLV backbone cleaved with SpeI and PacI (upper panel Figure 30). The plasmids extracted from the picked clones were digested with SpeI and PacI restriction enzymes. The expected digestion products of 5.9 kbp and 2.4 kbp were identified in all picked plasmids (lower panel Figure 30). In clones 1, 2, and 3, Sanger sequencing revealed that the insert sequences were in the correct orientation, however an A>G mutation was found at base 58 of the SgII sequence in all three SgII-GLuc plasmids. The base is marked in red in the sequencing alignments (Appendix, 6.4.8). This base change has no effect on the protein coding, so plasmid 3 was expanded, maxi-prepped and used for all relevant experiments.
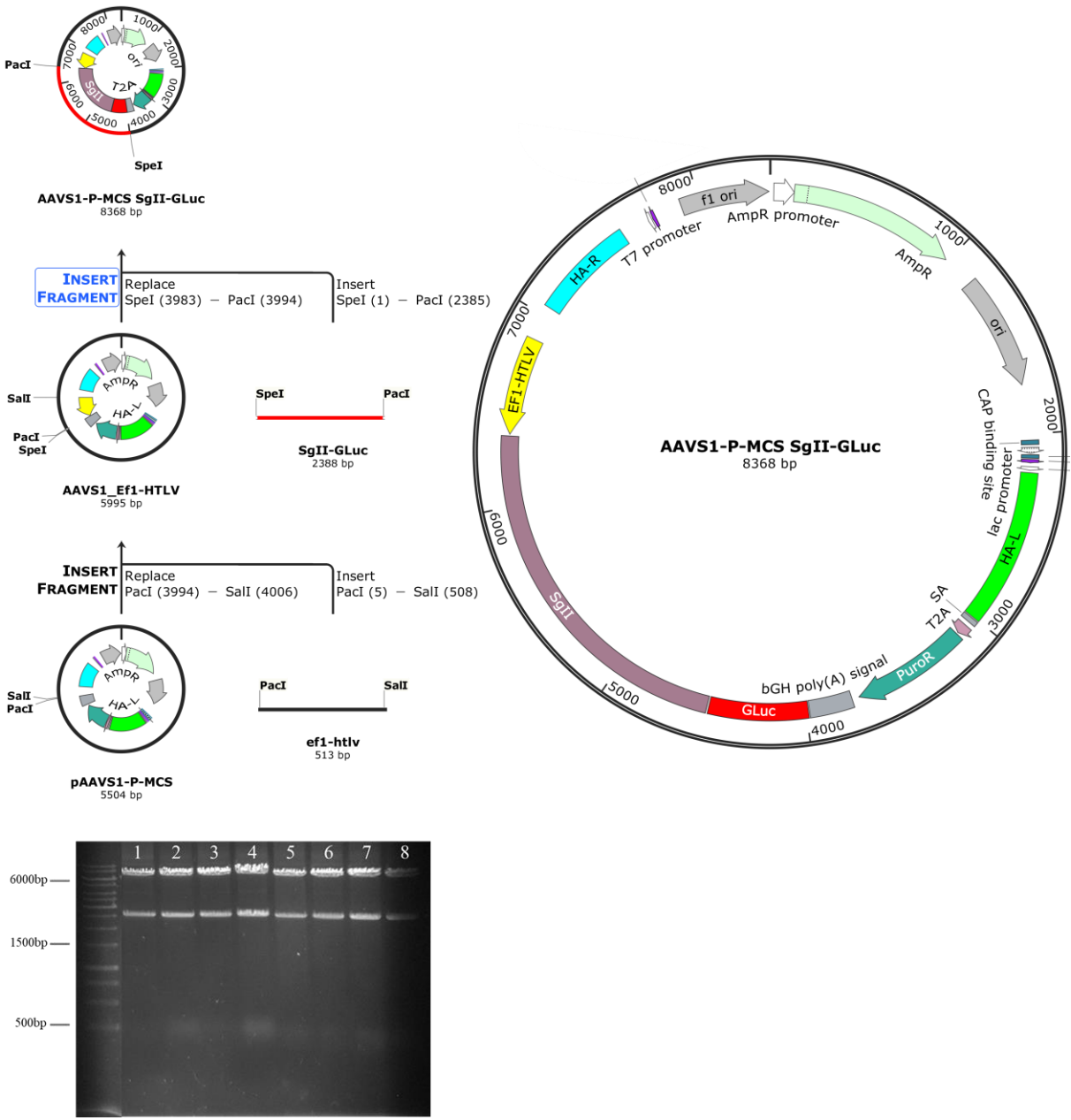
Section 3.1

VAMP2-GLuc donor plasmid

The VAMP2-GLuc donor plasmid contains the sequences coding for a fusion protein of VAMP2, a TEV protease recognition sequence, and no tag GLuc. The open reading frame of VAMP2 was amplified from OriGene plasmid RC207533[98] using flanking primers with 5' overhangs corresponding to PacI and KpnI. The TEV protease recognition sequence segment was constructed by annealing two inverse complementary oligonucleotide sequences with a 5' overhang imitating a digested KpnI restriction site and a 3' overhang imitating a digested EcoRI restriction site. The no tag GLuc sequence was amplified from the no tag GLuc donor plasmid described above using flanking primers with 5' overhangs corresponding to SpeI and EcoRI. The PCR products were digested with their corresponding restriction enzymes and the three segments were ligated together. The ligation reaction was amplified in a PCR reaction using the VAMP2 flanking primer with the 5' overhang for PacI and the GLuc flanking primer with the 5' overhang for SpeI. The PCR reaction was run on an agarose gel and the correctly sized PCR product of 900 bp was extracted. The VAMP2-TEV-GLuc fusion sequence was digested with the restriction enzymes SpeI and PacI and cloned into the pAAVS1-P-MCS Ef1-HTLV backbone cleaved with SpeI and PacI (upper panel Figure 31). The plasmid was extracted from the single clone resulting from this transformation and was digested with SpeI and PacI. The expected fragment sizes of 5.9 kbp and 900 bp were identified (bottom panel Figure 31). Sanger sequencing of this plasmid confirmed that all three components of the VAMP2-TEV-GLuc fusion were present and in the correct orientation with no mutations incorporated during the PCR (Appendix, 6.4.9). This plasmid was expanded, maxi-prepped, and used for all relevant experiments.
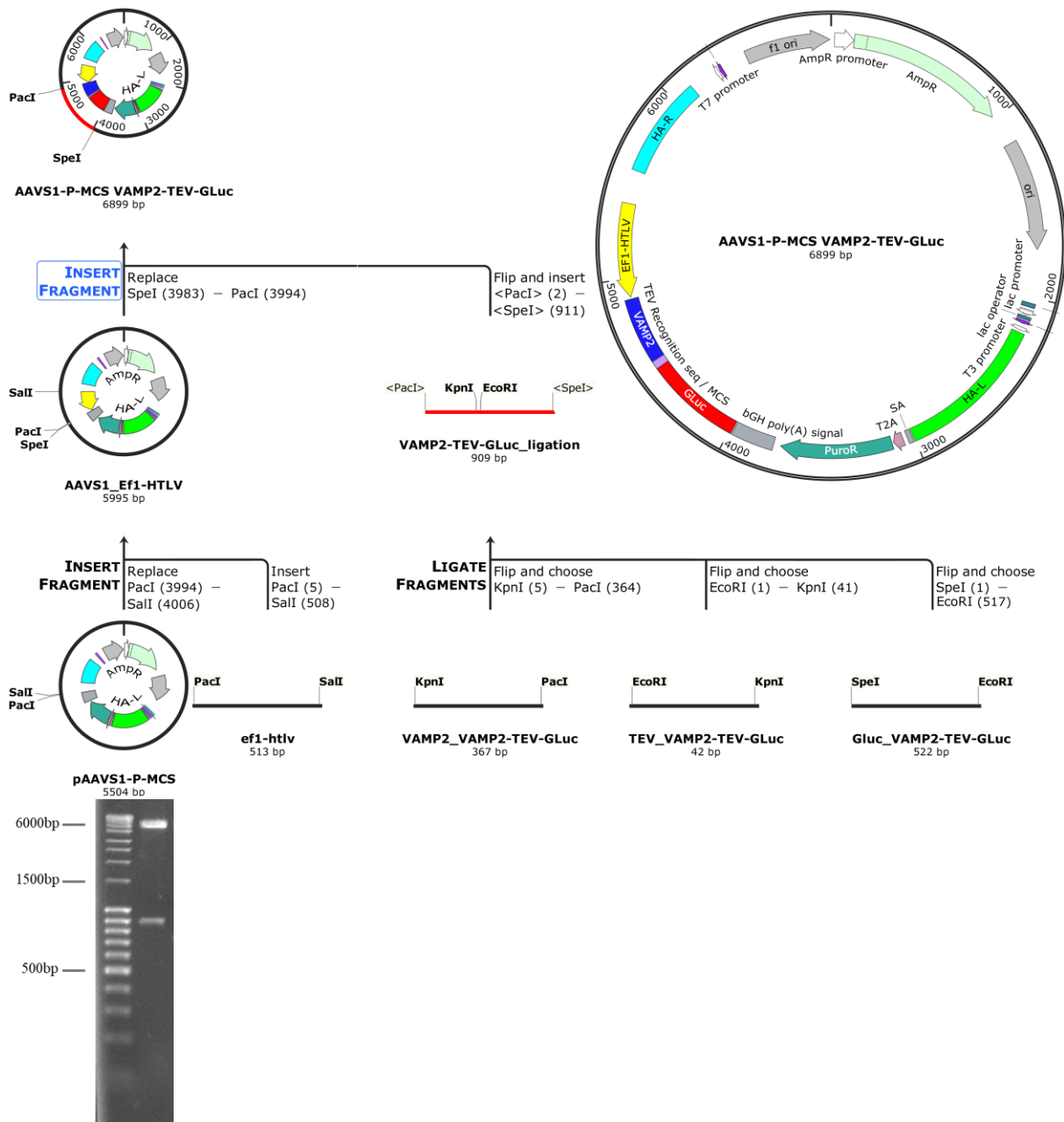
Figure 31: pAAVS1-P-MCS_Ef1-HTLV with VAMP2-GLuc Upper panel: History of the cloning method from pAAVS1-P-MCS to pAAVS1_Ef1-HTLV_VAMP2-TEV-GLuc. Lower panel: Plasmid map of the pAAVS1_Ef1-HTLV_VAMP2-TEV-GLuc donor plasmid and gel of digested mini-prepped clone. Expected product sizes of the digested plasmid with SpeI and PacI are 5.9 kbp and 900 bp.

3.1.3 Transfection with CRISPR/Cas9 and transient luciferase expression

The plasmids described above were used to co-transfect either SIMA neuroblastoma cells or IMR90-4 iPSCs. The eSpCas9(1.1) AAVS1_T2 plasmid was always part of the co-transfection, while the donor plasmid included in the transfection was dependent on the goal of the experiment and the desired sorting tag. The transient and stable transfection methods are described in section 2.4.

The transient transfection is an important first step to verify that the plasmid construction successfully induced GLuc expression. As mentioned above, SIMA cells transfected with the Ef1α hPOMC-GLuc plasmid did not express GLuc and therefore no luciferase activity was detected in the lysates, indicating a problem with the Ef1α promoter. On the other hand, the lysates of the SIMA cells transfected with the Ef1-HTLV hPOMC-GLuc plasmid contained high luciferase activity. All remaining donor plasmids described above with the Ef1-HTLV promoter could be used to produce clones transiently expressing GLuc (Figure 32). Luciferase levels in the lysate were only measured once. The differences in luciferase activity between each transfection was likely due to transfection efficiencies and not to the inherent properties of the constructs.
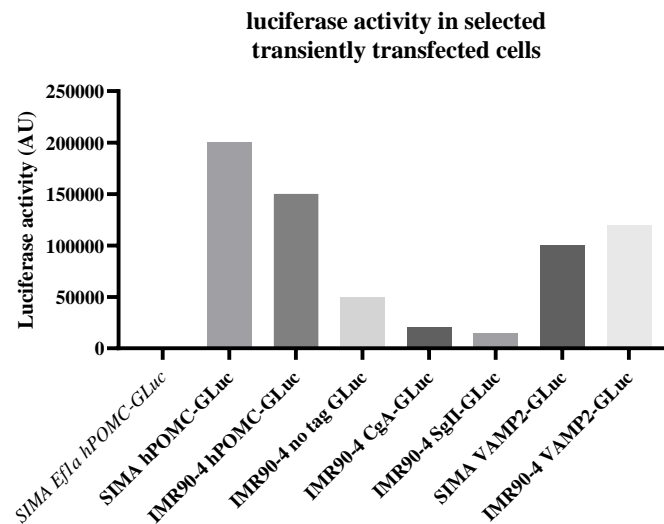
Figure 32: Luciferase activity in transiently transfected cells, n=1. SIMA/IMR90-4 cells were incubated with the transfection medium including transfection reagent and the appropriate donor DNA plasmid for 48 hours. Treated cells were collected, lysed, and their luciferase activity was measured by luminescence.

3.1.4 Post antibiotic selection – stable luciferase expression

Cells prepared for stable transfection were incubated post-transfection with the antibiotic puromycin to select those clones which had stably integrated the donor DNA, including the puromycin resistance gene, into their genome. The permanent expression of luciferase in the clones selected for the stable integration of the donor DNA in their gDNA is absolutely essential for the functionality of the MoN-Light BoNT assay. Therefore, as the clones with putative stable integrations of donor DNA divided and grew, it was necessary to confirm that these clones actually expressed luciferase by measuring its activity in cell lysates before proceeding. Cell lysates were collected when the clone reached 50-70 % confluency, whereby 70 % of the cells were split for clone expansion and the remaining 30 % of the cells were lysed and luciferase activity was measured. Almost all the isolated CRISPR-modified

SIMA hPOMC-GLuc clones expressed GLuc with luciferase activity levels ranging from several thousand AUs to almost 1,000,000 AUs. However, clones 26 and 31 apparently expressed little to no luciferase, as the activity measured in these lysates was similar to the blank measurement, indicating that the GLuc donor DNA may not have integrated into the gDNA of these clones (Figure 33).
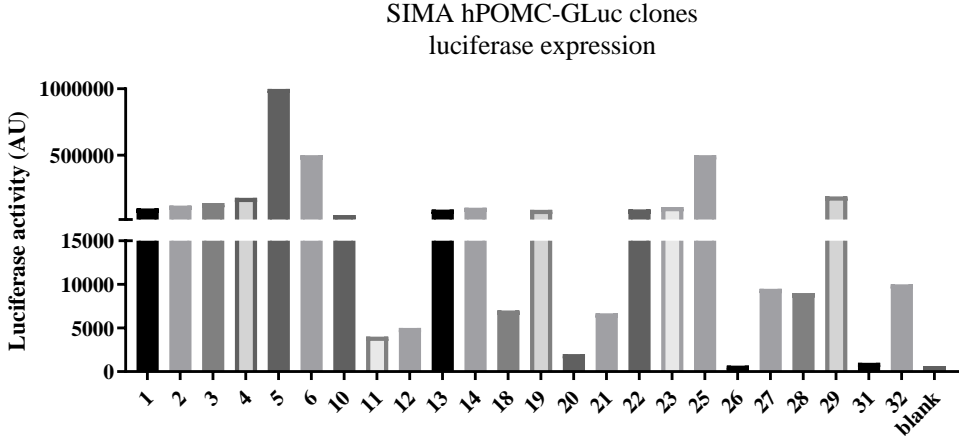
Figure 33: Luciferase activity in lysates of SIMA hPOMC-GLuc clones, n=1. SIMA cells were incubated with the transfection medium including Turbofect, the eSpCas9(1.1) AAVS1_T2, and the AAVS1-P-MCS hPOMC-GLuc donor plasmid for 48 hours. Post-puromycin selection, isolated cells were collected, lysed, and their luciferase activity was measured by luminescence.

Luciferase activity could also be measured in almost all the lysates of the IMR90-4 hPOMC-GLuc clones. The exception was clone 3, in which only a slightly higher luciferase activity could be detected than the blank. Otherwise, the luciferase activity measured from the remaining clones ranged from approximately 50,000 to 150,000 AUs (Figure 34).
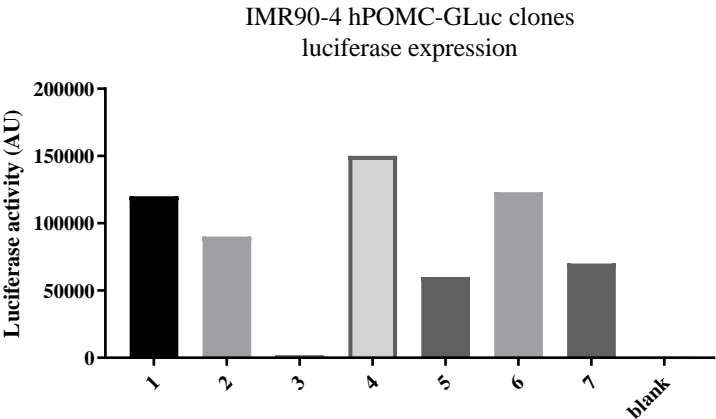
Figure 34: Luciferase activity in IMR90-4 hPOMC-GLuc clones, n = 1. IMR90-4 cells were incubated with the transfection medium including Lipofectamine3000, the eSpCas9(1.1) AAVS1_T2, and the AAVS1-P-MCS hPOMC-GLuc donor plasmid for 48 hours. Post-puromycin selection, isolated cells were collected, lysed, and their luciferase activity was measured by luminescence.

88

Twelve IMR09-4 no tag GLuc clones were successfully isolated from the stable transfection. Very low levels of luciferase activity were measured in the lysate of one clone, 11, with approximately 5,000 AU. The lysates from all the remaining clones had high measurements of luciferase activity, most far above 500,000 AU (Figure 35).
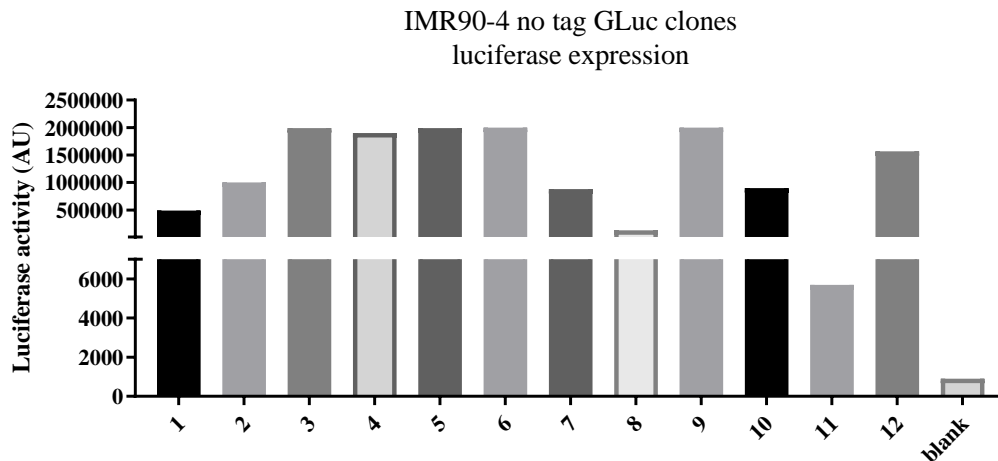
Figure 35: Luciferase activity in lysate of newly selected IMR90-4 no tag GLuc clones, n = 1. IMR90-4 cells were incubated with the transfection medium including Lipofectamine3000, the eSpCas9(1.1) AAVS1_T2, and the AAVS1-P-MCS no tag GLuc donor plasmid for 48 hours. Post-puromycin selection, isolated cells were collected, lysed, and their luciferase activity was measured by luminescence.

The luciferase activity in the lysates of the 16 IMR90-4 CgA-GLuc clones that survived antibiotic selection and expansion was generally lower than that detected in the hPOMC and no tag GLuc clones. The lysates of clones 2, 11, and 16 contained little to no luciferase, with detected activity levels comparable to the blank. However, almost all the remaining the clones measured luciferase activities ranging between 10,000 AU to 150,000 AU (Figure 36).
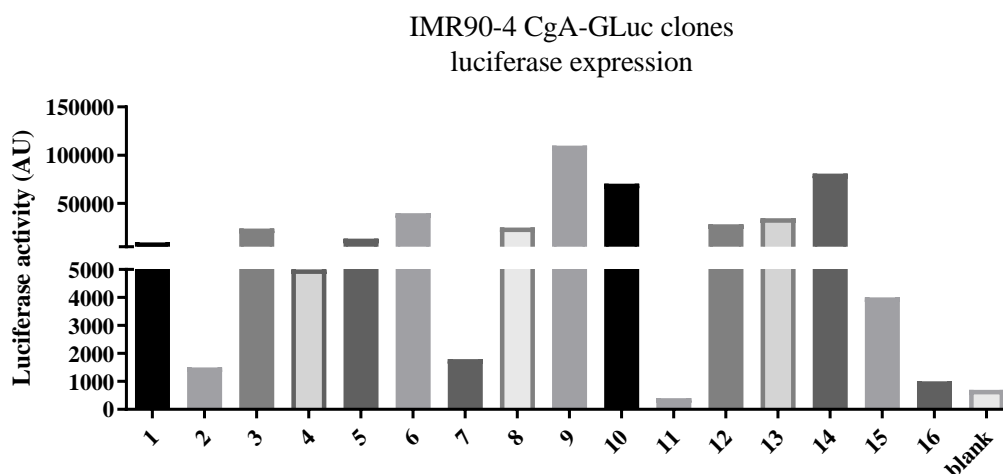
Figure 36: Luciferase activity in lysate of newly selected IMR90-4 CgA-GLuc clones, n = 1. IMR90-4 cells were incubated with the transfection medium including Lipofectamine3000, the eSpCas9(1.1) AAVS1_T2, and the AAVS1-P-MCS CgA-GLuc donor plasmid for 48 hours. Post-puromycin selection, isolated cells were collected, lysed, and their luciferase activity was measured by luminescence.

Section 3.1

Almost all of the lysates of the 11 isolated IMR90-4 SgII-GLuc clones contained only very low levels of luciferase activity. No activity could be measured in the lysate of the following clones: 5, 8, 11, 12, 13, 15, 16, 17, and 18. Only two clones, SgII-GLuc 4 and 10, had similar activity levels to the majority of CgA clones, between 10,000 AU and 150,000 AU. In the remaining clones only luciferase activity levels below 10,000 AU could be measured (Figure 37).
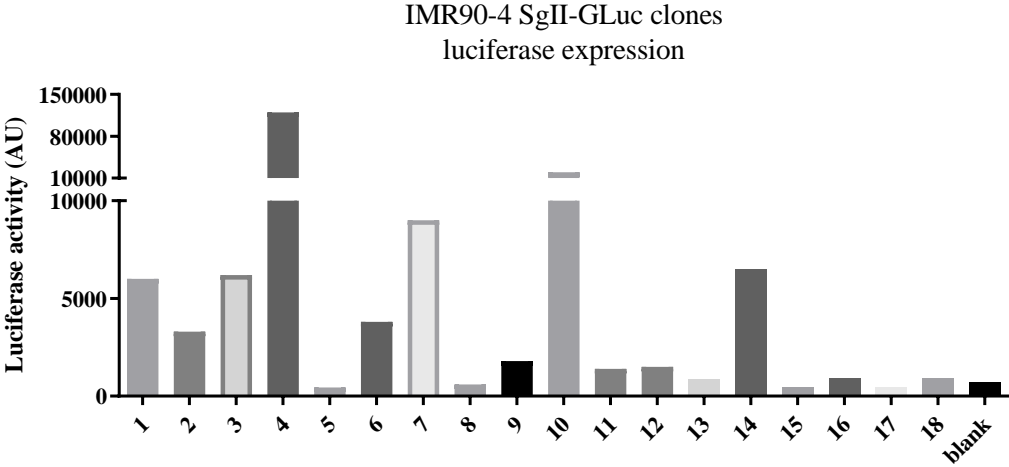
Figure 37: Luciferase activity in lysate of newly selected IMR90-4 SgII-GLuc clones, n = 1. IMR90-4 cells were incubated with the transfection medium including Lipofectamine3000, the eSpCas9(1.1) AAVS1_T2, and the AAVS1-P-MCS SgII-GLuc donor plasmid for 48 hours. Post-puromycin selection, isolated cells were collected, lysed, and their luciferase activity was measured by luminescence.

All CRISPR-modified SIMA VAMP2-GLuc clones isolated post-selection expressed luciferase, demonstrating low to medium levels of activity in the cell lysate ranging from 5000 AU to 15,000 AU (Figure 38).
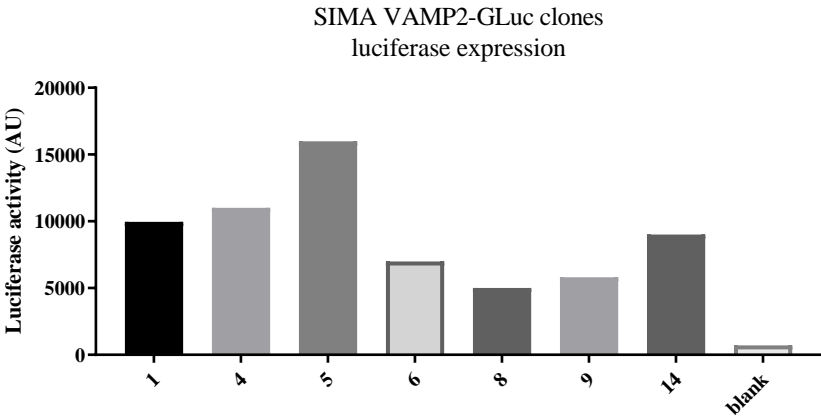
Figure 38: Luciferase activity in lysate of newly selected SIMA VAMP2-GLuc clones, n = 1. SIMA cells were incubated with the transfection medium including Turbofect, the eSpCas9(1.1) AAVS1_T2, and the AAVS1-P-MCS VAMP2-GLuc donor plasmid for 48 hours. Post-puromycin selection, isolated cells were collected, lysed, and their luciferase activity was measured by luminescence.

Section 3.1

Twenty-six IMR90-4 VAMP2-GLuc clones were isolated from the stable transfection. The activity of luciferase ranged markedly between these clones, from practically no activity in clones 10 and 18, to activity greater than 500,000 AU in clones 16 and 17. The majority of the clones contained medium activity levels ranging from 10,000 to 70,000 AU (Figure 39).
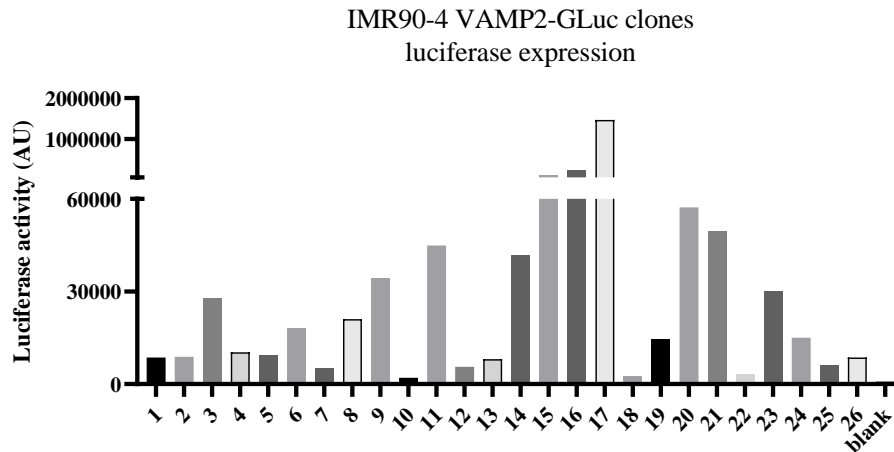
Figure 39: Luciferase activity in lysate of newly selected IMR90-4 VAMP2-GLuc clones, n = 1. IMR90-4 cells were incubated with the transfection medium including Lipofectamine3000, the eSpCas9(1.1) AAVS1_T2, and the AAVS1-P-MCS VAMP2-GLuc donor plasmid for 48 hours. Post-puromycin selection, isolated cells were collected, lysed, and their luciferase activity was measured by luminescence.

3.2   Establishment and validation of characterization techniques

In order to legitimize the use of any clone in the MoN-Light BoNT assay, it is necessary to fully characterize the cells on a genetic and proteomic level. The background and justification of the characterizations are extensively explained in the Introduction (section 1.6 – Goals and Strategies).

3.2.1   Confirmation of insert at AAVS1 safe harbor locus

Two sets of primers were initially used for the optimization of the amplification of the area surrounding the AAVS1 safe harbor locus. First, two primers each were designed to lay just 5' of the left homology arm (Fpcr_5'ofHA and 5'probeWT_F) and just 3' of the right homology arm (AAVS1xHA_R2 and AAVS1xHA_R3). All combinations of these forward and reverse primers should have amplified an approximately 1600 bp product in human WT gDNA. The second primer set was designated as Fpcr-803 and Rpcr-wt-183, which had been published as confirmation PCR primers for cloning in the AAVS1 safe harbor locus. These primers should have amplified a segment of DNA in non-transfected cells that is 1400 bp long[74]. In cells containing the donor DNA insert, the PCR product sizes should increase by a sequence length that depends on the donor plasmid (Table 53).

Donor DNA designation	Sequence length (bp)
no tag GLuc	2025
hPOMC-GLuc	2215
CgA-GLuc	3415
SgII-GLuc	3880
VAMP2-GLuc	2415

The Cas9 induced double strand break, the consequential integration of the donor DNA into the AAVS1 safe harbor locus by homology-driven recombination, and the locations of the first primers designed to amplify the genetic area surrounding the AAVS1 safe harbor locus are summarized in Figure 40.
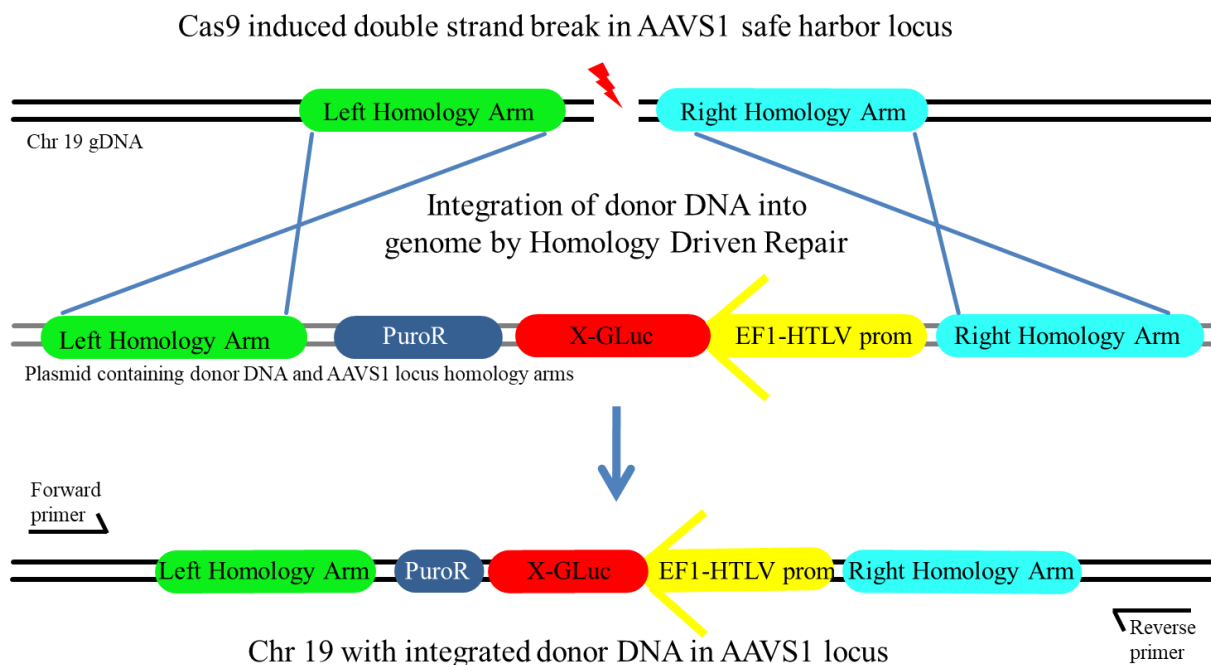
Figure 40: Sketch of integration of donor DNA into AAVS1 safe harbor locus by homology-driven recombination, including the positions of the first primers tested in order to verify correct insertion of the donor DNA. The "X-GLuc" refers to each of the GLuc constructs described in the plasmid production. Sequence length not to size.

None of the primer combinations using Fpcr_5'ofHA or 5'probeWT_F and AAVS1xHA_R2 or AAVS1xHA_R3 resulted in any amplicon from gDNA extracted from non-transfected cells. Nor did the purported positive control primers Fpcr-803 and Rpcr-wt-183 (see Appendix 6.5 for optimization details) result in any amplified DNA segments. The primer combination 5'probeWT_F, described above, with Rpcr-wt-3'HA, a primer located in the middle of the right homology arm, successfully amplified specific products with no additional contaminating non-specific bands at certain temperatures in a gradient using annealing

92

Section 3.2

temperatures ranging from 72 – 55 °C. The highest annealing temperatures produced WT allele specific products at 850 bp with the non-transfected SIMA gDNA (Figure 41 A). In a temperature gradient at the same annealing temperature range run using SIMA hPOMC-GLuc clone 5 gDNA, an additional band can be seen at 3000 bp, exactly the size of the expected product including the donor DNA (Figure 41 B). Interestingly, the WT band at 850 bp still amplifies in the clone DNA. This indicates that there is probably one WT allele in this clone as well as one allele in which the donor DNA has successfully integrated, making it a heterozygous clone.
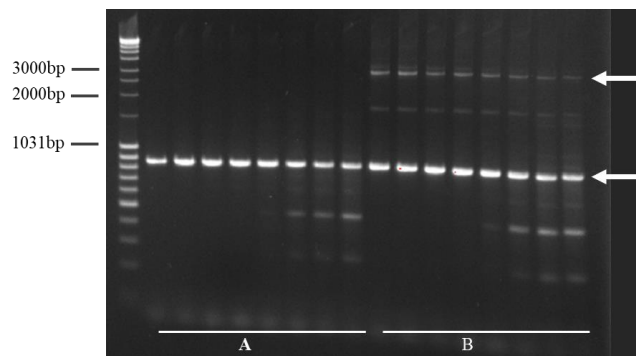
Figure 41: Successful amplification of the AAVS1 safe harbor locus by temperature gradient PCR (72 – 55 °C). (A) SIMA non-transfected gDNA, (B) SIMA hPOMC-GLuc clone 5 gDNA. Expected product sizes indicated by arrows: WT 850 bp, Insert 3000 bp.

The final insert confirmation amplicon spans a sequence area defined by the forward primer 5'probeWT_F, found 5' of the left homology arm, and the reverse primer Rpcr-wt-3'HA, found in the middle of the right homology arm (Figure 42). If a successful insertion into the AAVS1 safe harbor locus has taken place, the entire donor DNA sequence is amplified by these primer pairs.

Chr 19 with integrated donor DNA in AAVS1 locus
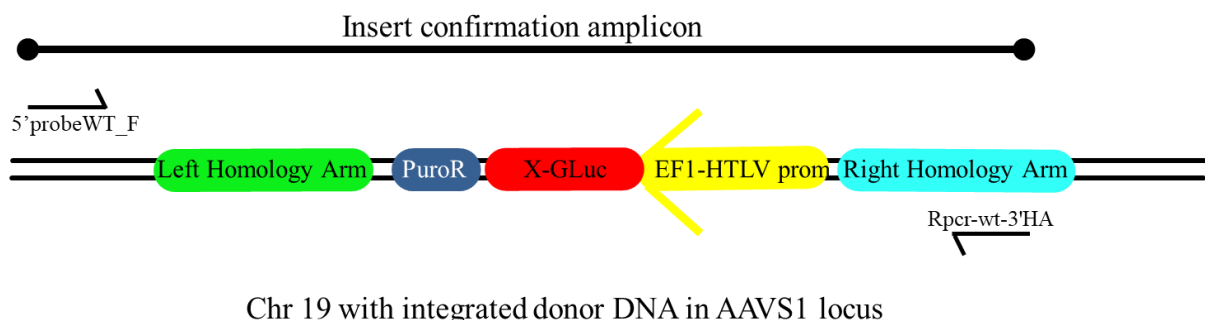
Figure 42: Illustration of the final locations of the primers used to confirm the integration of donor DNA into the AAVS1 safe harbor locus. The "X-GLuc" refers to all GLuc constructs described in the plasmid production. Sequence length not to size.

3.2.2 Analysis of possible off-target integration events

3.2.2.1 Southern blot

Southern blot was tested to identify possible off-target donor DNA integrations using four different restriction enzymes. Included in the optimization were gDNA from a selection of CRISPR-modified SIMA hPOMC-GLuc clones, as well as gDNA from the Random_Insertion-hPOMC1-26GLuc prototype clone[2] (see Introduction, section 1 for prototype description) which was the basis for the development of the MoN-Light BoNT assay. The integration site of the donor DNA at the AAVS1 safe harbor locus in the hPOMC-GLuc clones is known, since this integration was directed by the CRISPR/Cas9 genetic editing method and the successful integration can be verified by the insert confirmation PCR technique described in section 3.2.1. Recognition sites of each restriction enzyme used to test the Southern blot technique can be localized in the region flanking the AAVS1 safe harbor locus, and the expected DNA fragment sizes which should include the donor DNA can be calculated (Table 54). The expected digested DNA fragment sizes for the Random_Insertion-hPOMC1-26GLuc prototype clone gDNA cannot be predicted, since the integration of GLuc in the genome was random.

Table 54: Summary of restriction enzymes and expected fragment sizes for hPOMC-GLuc clones in Southern blot off-target donor DNA analysis.

Restriction enzyme	Expected DNA fragment size
VspI	36 kbp
HindIII	1.8 kbp
BglI	1.5 kbp
EcoRI	10.3 kbp

Optimization and troubleshooting

The standard Southern blot protocol from Roche (described in Materials and Methods section 2.9) was carried out with the CRISPR/Cas9 modified SIMA hPOMC-GLuc clones and the Random_Insertion-hPOMC1-26GLuc prototype clone. Very faint bands could be detected in one exemplary SIMA hPOMC-GLuc clone at the correct sizes under specific hybridization and washing conditions. However, the faint bands were not identifiable in all SIMA hPOMC-GLuc clones tested, despite the fact that the insert was proven by PCR to have been incorporated at the AAVS1 safe harbor locus. Furthermore, no bands were detected in the Random_Insertion-hPOMC1-26GLuc prototype clone, despite the known stable incorporation of GLuc into a random genomic locus in this clone. The purpose of the Southern blot is to exclude any off-target insertions, therefore in order to meet the criteria of a reliable

94

characterization technique, the method must be able to identify all known and random insertions and may not produce any false negative results. Due to the unreliable signal yield using the standard protocol, a series of experiments were done to optimize the Southern blot protocol.

Various factors can be changed in the Southern blot optimization, including the size of the probe, which unique element of the donor DNA the probe detects (GLuc, the puromycin resistance sequence, part of the non-human promoter), the hybridization temperature of the probe, and the temperature of the high stringency wash. All optimization protocols tested are summarized in Table 55. It is important to note that the random insertion of the Gaussia luciferase in the Random_Insertion-hPOMC1-26GLuc prototype clone will only be detected with a GLuc probe, since this is the only element in the donor DNA sequence that matches the insertion into the SIMA Random_Insertion-hPOMC1-26GLuc prototype clone sequence. None of the probes or hybridization conditions resulted in reproducible and strong bands and no condition allowed the identification of GLuc in the Random_Insertion-hPOMC1-26GLuc clone genome. The main criteria for the off-target integration test was not met: the method could not identify all known inserts and continued to produce false negative results. Therefore the Southern blot method was rejected (see Appendix, 6.6.1 for representative blots).

Table 55: Summary of southern blot probes and tested hybridization temperatures. Deviation from standard protocol: # = High stringency temperature at 60 °C; § = Detection solution CDP Star; & = Digested samples run on Roche DIG Membrane; % = Hybridization with 3µl/ml probe.

Probe	Probe Size (bp)	Hybridization Temperature(s) (°C)
GLuc F1R1	510	45, 50, 55, 60
GLuc F2R2	395	44, 47, 52, 57[&]
GLuc F2R4	221	44, 50, 53, 57, 57[#], 57[§], 57[&], 57[%]
GLuc F3R3	145	49
GLuc F3R4	299	50
Puromycin	382	58
EF1α-HTLV	430	50, 53, 55
HTLV	266	50, 50[§]

3.2.2.2 Ligation-mediated PCR

As described in the Materials and Methods section 2.10, three oligonucleotide adapters were designed to match the sticky-ends of DNA which had been digested by the restriction enzymes HindIII, BspHI, or AseI. The gDNA of a SIMA hPOMC-GLuc clone was digested with either HindIII, BspHI, or AseI and then ligated to the corresponding annealed adapter oligonucleotides. The ligated products were then used as the DNA template in a PCR reaction with a forward primer found in the GLuc coding sequence and with the antisense adapter

oligonucleotide as the reverse primer. A standard temperature gradient PCR protocol was run in order to find any possible annealing temperatures which might correctly amplify any product from the ligation reaction. Since the insertion confirmation test was already carried out for the samples tested and it was confirmed that there was at least one copy of donor DNA inserted at the AAVS1 safe harbor locus, there should at least be one correct product, in addition to any off-target products. None of the potential products were successfully amplified. Because none of the initial protocols yielded the expected PCR product that should have been amplified from the confirmed correct insert, it appears that even with extensive optimization, the assay cannot be rendered sufficiently sensitive to exclude potential false negative results. Because it is essential that no off-target inserts are missed by the assay, the approach was no longer pursued (see Appendix, 6.6.2 for more details).

3.2.2.3   Establishment of technique for copy number quantification using qPCR analysis

An alternative approach to identify potential off target inserts is quantitative PCR to determine the copy number of a specific sequence. In order to calibrate this PCR, endogenous genes that are coded by autosomes (n = 2) and X-Chromosomes (ChrX, n = 1 in male, n = 2 in female) were amplified and the threshold cycles were compared. First, it was confirmed that the qPCR method could detect small changes in the gDNA concentration, which was analyzed with a doubling experiment in which two samples, one each from a female and male, were amplified with two different primer sets found on ChrX covering a segment of the genes *GATA1* and *TMBS15B*. Three quantities of gDNA were amplified: 10 ng, 20 ng, and 40 ng. The fold difference between sample amplification was calculated by ΔCt, whereby the smallest amount of gDNA was used as the reference gene. It can be seen that in both female and male samples, the qPCR method was sensitive enough to detect double and quadruple the amount of DNA added to the reaction (Figure 43). Importantly, however, despite the fact the male samples should have half the number of copies of the ChrX genes, it is impossible to detect this difference between samples, since the amount of DNA and hence the number of copies of the DNA in the sample are normalized to the first concentration added to the reaction. In this case, whether the starting copy number is one, or two, or 18, double the amount of DNA in the reaction will always result in a 2-fold increase of signal in the qPCR. Therefore, an additional normalization step must be added in order to determine the amount of starting material for each specific sample.

96
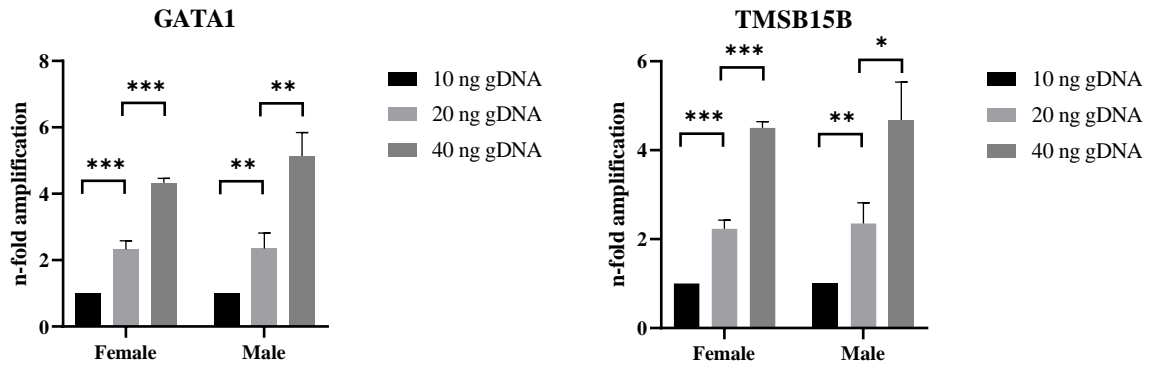
Figure 43: gDNA doubling and quadrupling experiment with female and male samples. Amplification of 10 ng, 20 ng, and 40 ng gDNA, doubling calculation by ΔCt. Values displayed mean ± SD, statistical differences are determined with the two-tailed, unpaired t-test. n = 3, * = p < 0.05, ** = p < 0.01, *** = p < 0.001.

The normalization problem was solved by comparing two genes from the same sample to one another, one of which resides on an autosome, which has a copy number of two, while the other gene resides on a sex chromosome, which will have a copy number of two in females or a copy number of one in males. The copy number of four ChrX genes (*RBBP7, GATA1, HPRT1, TMSB15B*) was compared to the copy number of one autosomal gene, *CHOP*. The copy number of all four of the ChrX genes was twice as high in female samples compared to the male samples. The samples were normalized by taking the quotient of the ChrX to CHOP, whereby in the graphic, $1.0 = 2$ copies (autosome vs allosome in females) and $0.5 = 1$ copy (autosome vs allosome in males) (Figure 44). This test indicates that the qPCR method should be sensitive enough to detect a single change in the copy number of a gene in gDNA samples.
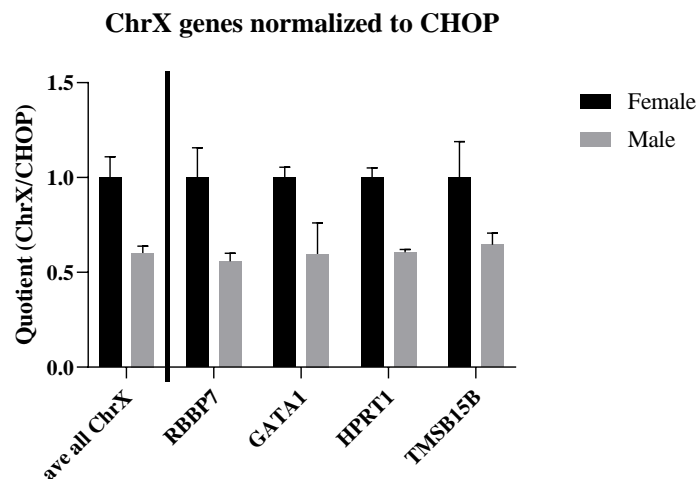
Figure 44: qPCR analysis of ChrX genes in female and male samples. Change in copy number calculated by ΔCt in ChrX normalized to autosomal gene CHOP. Values displayed mean ± SD, n = 3.

The normalization of the copy number and concentration are important factors to consider in the attempt to quantify gene copy numbers, however, when using and comparing

multiple primer pairs to amplify various genetic segments, it is also important to take into account possible differences in primer efficiencies. For this reason, one more control step was added to this quantification technique, which can be used to normalize all three of these disrupting factors in the qPCR. Specifically, a plasmid was constructed to contain the autosomal gene sequence for CHOP, the genetic sequence for one of the ChrX genes (either *RBBP7*, *GATA1*, *HPRT1*, or *TMSB15B*), and the sequence for GLuc. This plasmid can be included and amplified with each primer pair, alongside any clone gDNA for normalization and control purposes. Both the copy number of the ChrX gene and GLuc can be determined and then compared to one another for verification purposes (Figure 45).

Plasmid: 1x Autosomal Reference Gene sequence and 1x ChrX sequence and 1x GLuc sequence
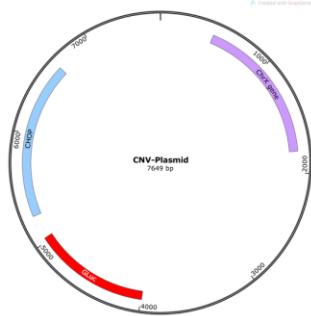
Amplify gDNA dilution and Plasmid Z dilution with autosomal and ChrX reference gene and GLuc primers
→ ref gene: Plasmid, ct=25; gDNA, ct=25; n=2 and ($2^{25-25}= 1$)
→ ChrX: Plasmid, ct=a; whereby a=n=2
　　→ (a might not equal 25, due to differences in primer efficiency)
　　→ $2^{a-ct\_gDNA}= b$
→ GLuc: Plasmid, ct=c; whereby a=n=2
　　→ (a might not equal 25, due to differences in primer efficiency)
　　→ $2^{c-ct\_gDNA}= d$
　　　　→ if d = 0.5, then n=1 (single insert, heterozygote)
　　　　→ if d = 1, then n=2 (homozygote)
　　　　→ if d > 1, then multiple inserts
→ Compare b and d to ensure calcultions meet expectations
　　→ if b/d in female cells = 1, then n = 2

*Figure 45: Summary of control plasmid for copy number verification. *genetic elements in plasmid not to size**

The ChrX and GLuc copy number values were analyzed and compared according to this method. Male and female gDNA served as controls. The copy number for both ChrX genes in the control female was approximately 2, while the copy number for both genes in the control male was approximately 1. Through the rest of the diagram the two ChrX gene copy numbers depicted in greyscale can be compared to the GLuc copy numbers colored in red. SIMA cells are derived from a male donor and only have one copy of ChrX, while the IMR90-4 cells are derived from a female donor and have two copies of ChrX. It can also be seen that the amplification of the ChrX genes is consistent both between samples and between genes (Figure 46).

Section 3.2
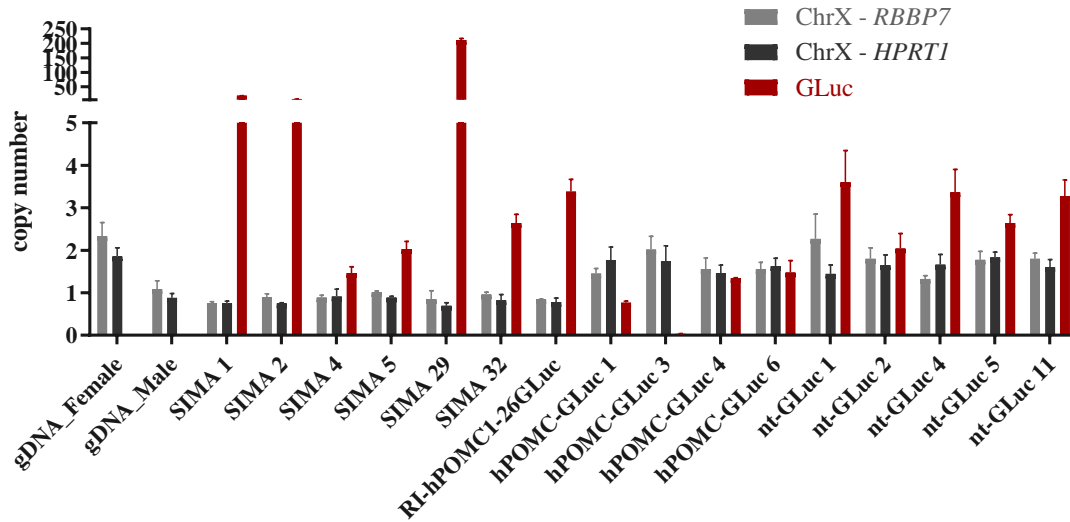
Figure 46: Copy numbers of RBBP7, HPRT1, and GLuc in selected clones. Copy number calculated by ΔCt of GLuc/ChrX normalized to corresponding plasmid amplification and CHOP amplification. RI-hPOMC1-26GLuc is short for Random_Insertion-hPOMC1-26GLuc, nt-GLuc is short for no tag GLuc. Values displayed mean ± SD, n = 3.

In order to optimize the graphical representation of this data and its ease in interpretation, the ratio of GLuc to ChrX gene was calculated and displayed as a single bar in the graph (Figure 47 A). Furthermore, the ratio of RBBP7 to CHOP in control female and male will always be included in order to confirm the quality of the set of data (Figure 47 B). The analyses of GLuc copy number in the following sections will be presented in this style. Due to the use of two control amplifications (the autosomal and ChrX), this test to characterize CRISPR/Cas9 donor DNA insertion numbers is termed the double-control quantitative copy number PCR (dc-qcnPCR).
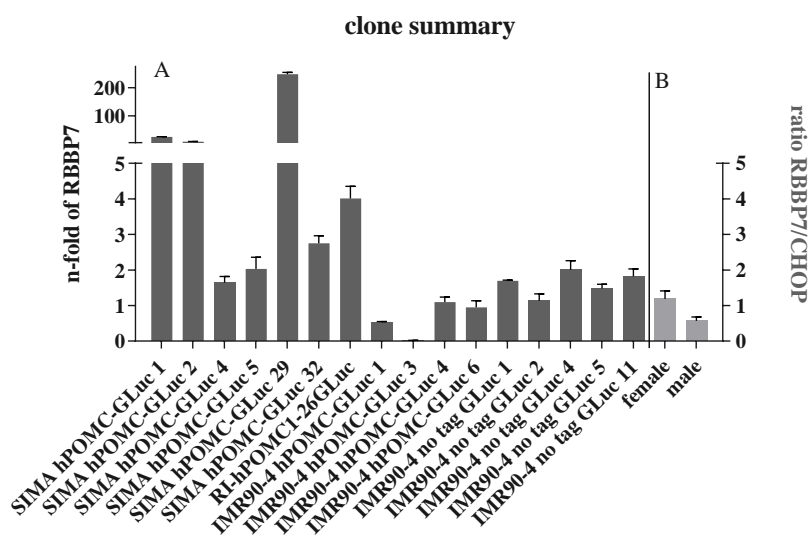
Figure 47: Ratio of GLuc to ChrX gene RBBP7. A: Copy number calculated by ΔCt of GLuc normalized to corresponding plasmid amplification, CHOP and ChrX amplification. B: Ratio of RBBP7 to CHOP in female and male gDNA. RI-hPOMC1-26GLuc is short for Random_Insertion-hPOMC1-26GLuc. Values displayed mean ± SD, n = 3.

99

3.2.3 Cellular localization of Gaussia Luciferase

3.2.3.1 Differential fractionation

The use of differential fractionation to identify the localization of a protein takes advantage of the ability to use multiple centrifugation steps to separate homogenized cells into fractions containing cellular debris, cytosolic proteins, and membrane proteins. After each centrifugation step, the cellular components were collected and resuspended and the activity of luciferase was measured. Contrary to expectations, the highest amount of luciferase was by far found in both the cellular debris and the cytosolic fractions, with only minimal luciferase activity detected in the small vesicle fraction (specific data can be found in the Appendix, section 6.6.3.1). The most likely reason for this unexpected distribution is that the intracellular vesicles that contained the luciferase were transiently or permanently broken by the homogenization procedure. Since no adequate control was available to check for the integrity of the vesicles after homogenization, this method had to be abandoned.

3.2.3.2 Immunofluorescence Primary Antibody Validation

The second method tested to determine the correct sorting of GLuc into secretory vesicles is immunofluorescence (IF). In order to accurately carry out double staining of cellular proteins by IF, it is necessary to verify that the antibodies specifically label the protein of interest. The following primary antibodies were used for the first time in this lab for immunofluorescence:

- Gaussia luciferase (rabbit)
  - New England Biolabs (NEB)
  - Santa Cruz (SC)
- Golgi97 (mouse)
- Lectin HPA (Alex-fluor coupled Golgi marker)
- GM130 (mouse)
- Chromogranin A (CgA) (goat)
- Secretogranin II (SgII) (rabbit)
- Chromogranin A (CgA) (mouse)
- Secretogranin II (SgII) (mouse)
- Synaptophysin (Syp) (mouse)
- GAPDH (mouse)
- Islet1 [1B1] (mouse)
- Islet1 [1H9] (mouse)

Gaussia luciferase

The NEB rabbit-α-GLuc primary antibody was tested in an IMR90-4 hPOMC-GLuc clone as well as with non-transfected IMR90-4 cells, to prove the specificity of both the NEB GLuc primary antibody and the secondary antibody (Alexa-fluor anti-rabbit 488). Two dilutions were prepared for both cell types: 1:500 (Appendix 6.6.3.2, Figure 141) and 1:1000. Both dilutions labeled the GLuc protein well, but the 1:1000 dilution produced less background staining in the non-transfected cells, therefore all subsequent experiments were performed using the 1:1000 dilution of the NEB rabbit-α-GLuc antibody (Figure 48).
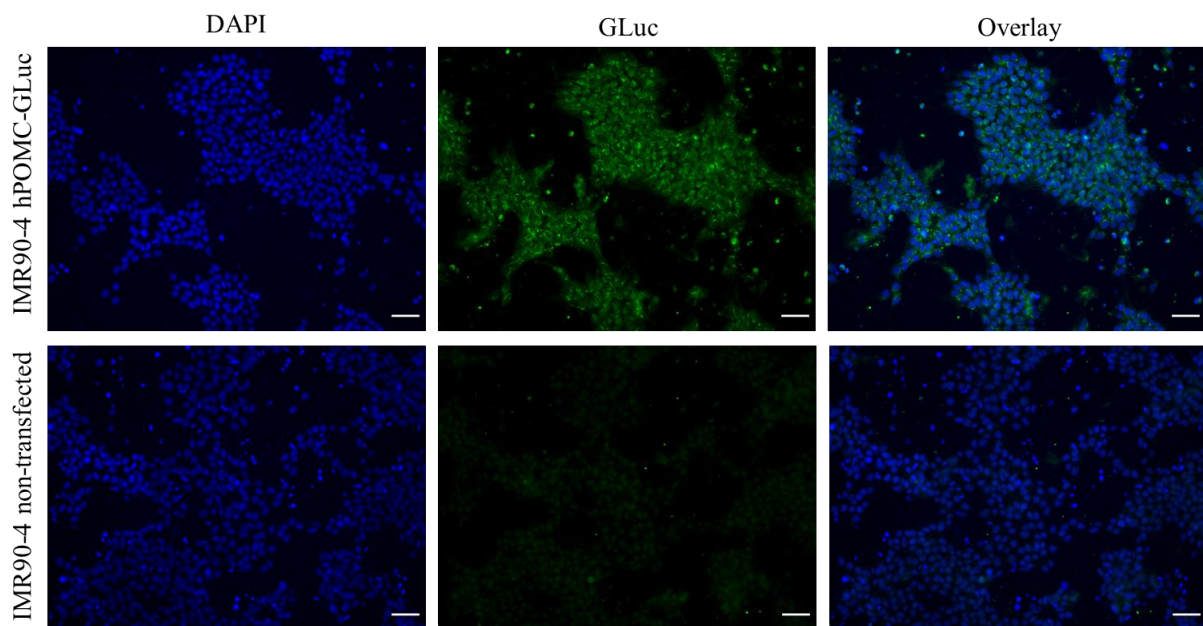
Figure 48: IMR90-4 hPOMC-GLuc and non-transfected cells stained with IF indirect labelling method. Cells were fixed, permeabilized and then incubated overnight with a 1:1000 dilution of rabbit-α-GLuc primary antibody. Secondary antibody incubation with Alexa-fluor α-rabbit 488 diluted to 1:750. Cells were additionally incubated with DAPI for nuclei staining. scale bar = 50μm

The SC rabbit-α-GLuc antibody was tested at a 1:50 dilution in IMR90-4 hPOMC-GLuc cells, but failed to stain (Appendix 6.6.3.2, Figure 142). This antibody was not used any further in this project.

Golgi apparatus

Golgin97 is a marker for the *trans*-Golgi network. The mouse-α-Golgin97 antibody was tested with multiple dilutions (1:50, 1:500, 1:750, and 1:1000) and with 2 different secondary antibodies (anti-mouse Alexa-fluor 532 and anti-mouse Alexa-fluor 680) in SIMA, IMR90-4, and HepG2 cells. None of the dilutions with either of the secondary antibodies resulted in successful staining of the *trans*-Golgi network. A replacement Golgi identification marker, Lectin Helix pomatia agglutinin (HPA) was tested, which selectively binds to α-N-

acetylgalactosamine residues, an intermediate sugar found on serine and threonine residues transferring between the *cis*-Golgi and the *trans*-Golgi. The Lectin HPA was conjugated to Alexa-fluor 647 and did not need to undergo secondary antibody incubation. Under various permeabilization conditions, the Lectin HPA marker also did not stain the Golgi (more details in the Appendix 6.6.3.2). The Golgin97 antibody was tested with an optimized protocol (see Materials and Methods 2.14.1). The alternate method resulted in only faint staining of HepG2 and SIMA cells (Appendix 6.6.3.2).

Because of the suboptimal performance of the Golgin97 and Lectin HPA markers up to this point, a primary antibody for the *cis*-Golgi, mouse-α-GM130 was tested. This antibody immediately successfully stained the Golgi apparatus illuminating the typical accumulation of the organelle in the perinuclear centrosomal region[110] (Figure 49).
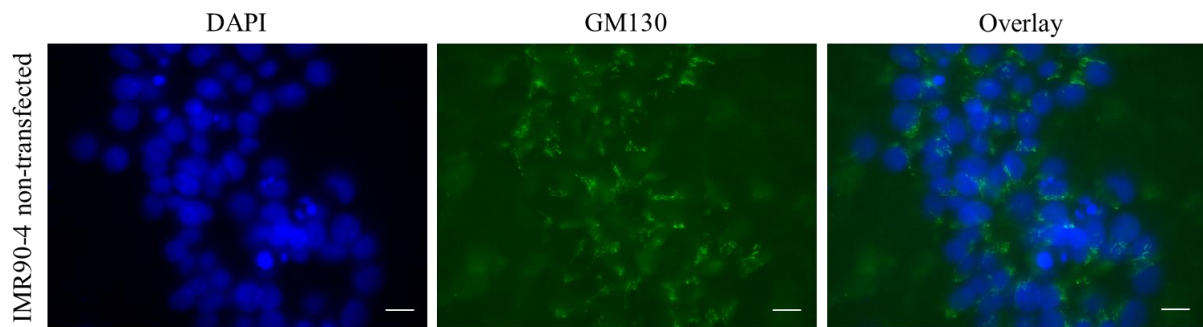
Figure 49: Non-transfected IMR90-4 cells stained with IF indirect labelling method. Cells were fixed, permeabilized and then incubated overnight with a 1:250 dilution of mouse-α-GM130 primary antibody. Secondary antibody incubation with Alexa-Fluor α-mouse 488 diluted to 1:1500. Cells were additionally incubated with DAPI for nuclei staining. scale bar = 10μm

Large dense core vesicles

The 1:50 dilution of goat-α-CgA successfully labelled the axonal outgrowths of differentiated CRISPR-modified SIMA cells, but did not result in any staining in CRISPR-modified hPOMC-GLuc motor neurons differentiated according to Maury *et al* for 30 days (Figure 50).

102
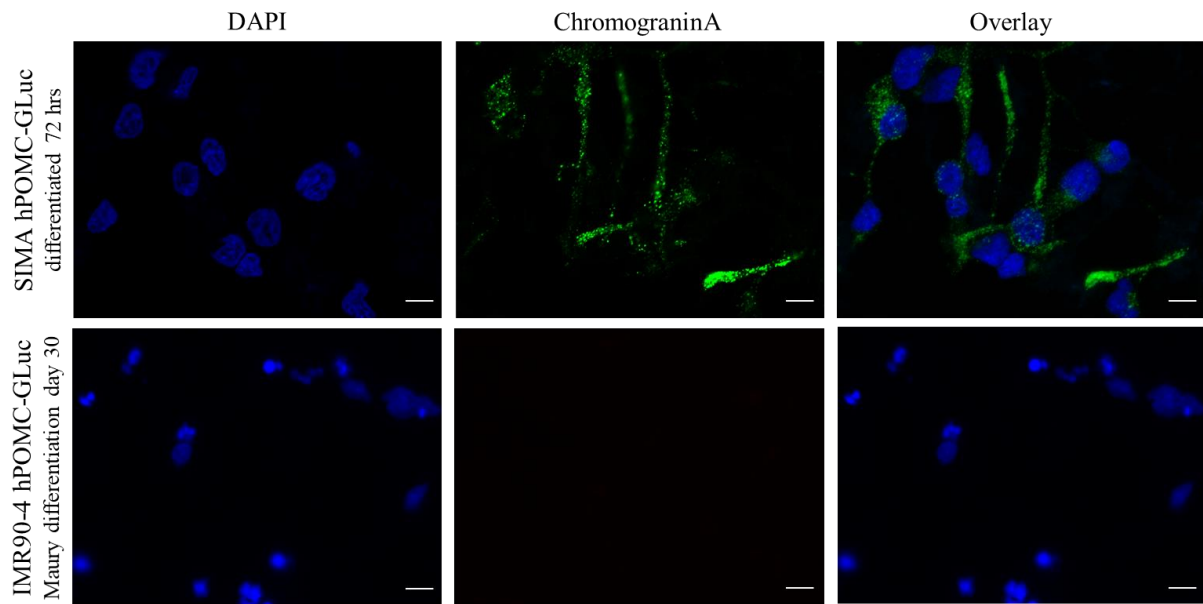
Figure 50: Differentiated SIMA and IMR90-4 hPOMC-GLuc (Maury Day30) clones stained with IF indirect labelling method. Cells were fixed, permeabilized and then incubated overnight with a 1:50 dilution of goat-α-CgA primary antibody. Secondary antibody incubation with Alexa-Fluor α-goat 488 diluted to 1:1500. Cells were additionally incubated with DAPI for nuclei staining. scale bar = 10µm

The rabbit-α-SgII antibody, successfully labelled differentiated CRISPR-modified SIMA cells both with the 1:1000 dilution (top panels) as well as with the 1:250 dilution (bottom panels). The protein can be found in both the axonal outgrowths and congregated around the perinuclear centrosomal region (Figure 51).
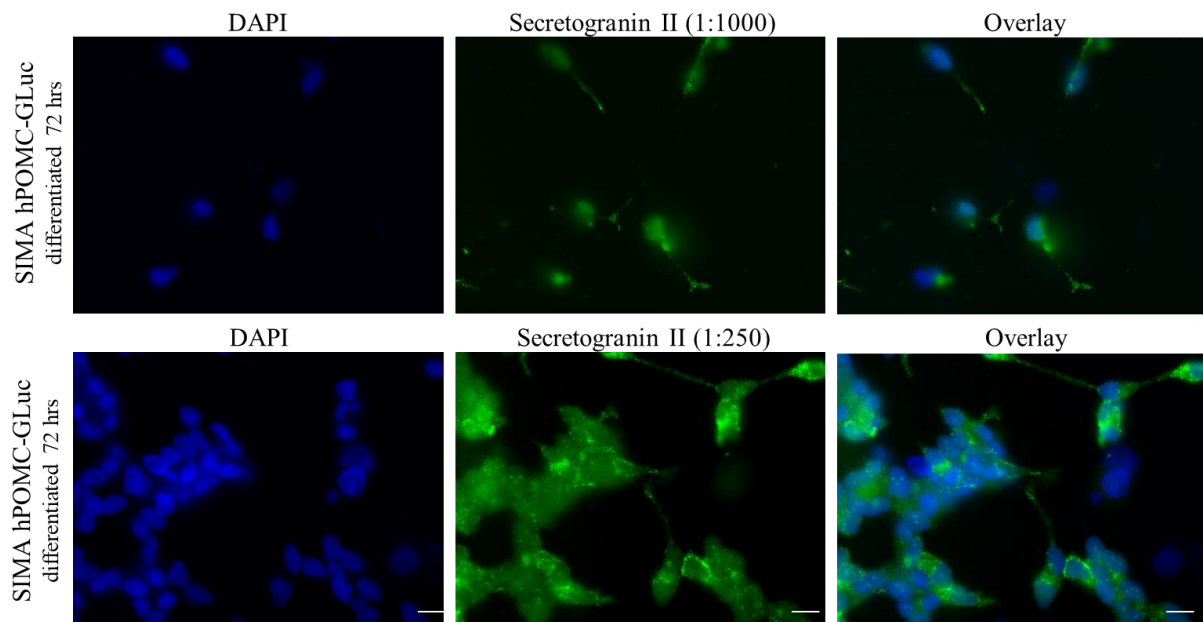
Figure 51: Differentiated SIMA hPOMC-GLuc clone stained with IF indirect labelling method. Cells were fixed, permeabilized and then incubated overnight with a 1:1000 and 1:250 dilutions of rabbit-α-SgII primary antibody. Secondary antibody incubation with Alexa-Fluor α-rabbit 488 diluted to 1:1500. Cells were additionally incubated with DAPI for nuclei staining. scale bar = 10µm

The mouse-α-CgA antibody was tested to determine if it could successfully label differentiated motor neurons (Figure 52). The manufacturer's recommended dilution was 1:100, and this was tested in differentiated SIMA hPOMC-GLuc cells (panels A-C) and in CRISPR-modified motor neurons differentiated according to Maury *et al* for 30 days (panels D-F). Similar to the goat-α-CgA antibody (Figure 50), CgA was successfully labeled in the SIMA cells, but not in the MNs, indicating that the problem was not with the original antibody, but more likely due to a possible lack of CgA in MNs.
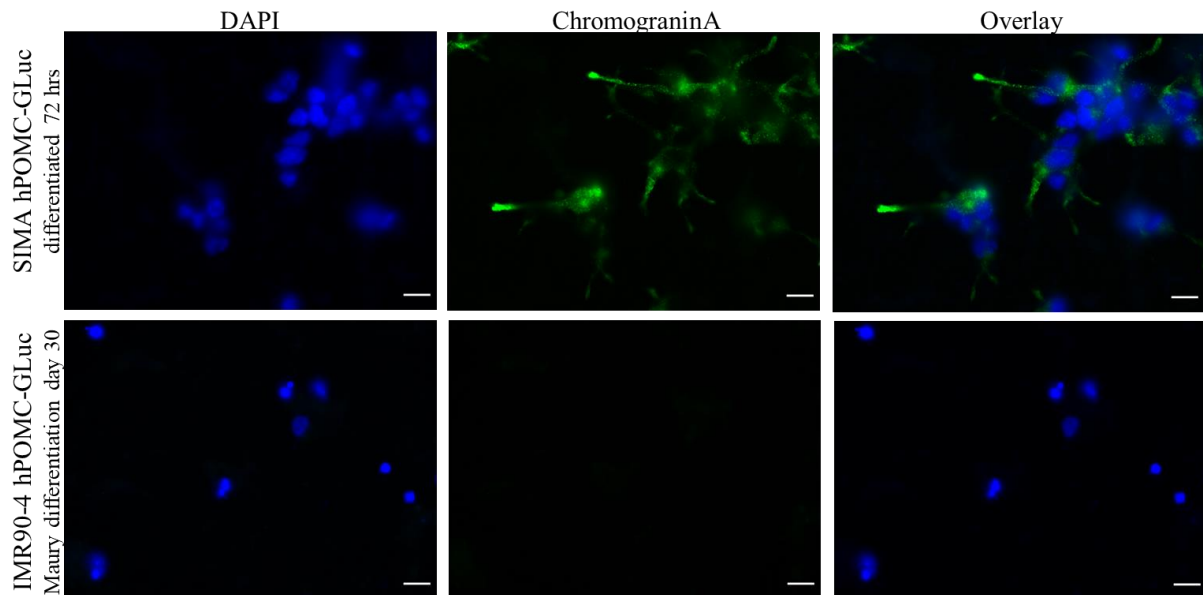
Figure 52: Differentiated SIMA hPOMC-GLuc clone and a IMR90-4 hPOMC-GLuc clone differentiated according to Maury et al stained with IF indirect labelling method. Cells were fixed, permeabilized and then incubated overnight with a 1:100 dilution of mouse-α-CgA primary antibody. Secondary antibody incubation with Alexa-Fluor α-mouse 488 diluted to 1:1500. Cells were additionally incubated with DAPI for nuclei staining. scale bar = 10μm

The mouse-α-SgII antibody was tested according to the manufacturer's suggestion to 1:500, and successfully labelled both differentiated SIMA hPOMC-GLuc cells (top panel) and CRISPR-modified motor neurons differentiated according to Maury *et al* for 30 days (bottom panel Figure 53).

	DAPI	Secretogranin II	Overlay
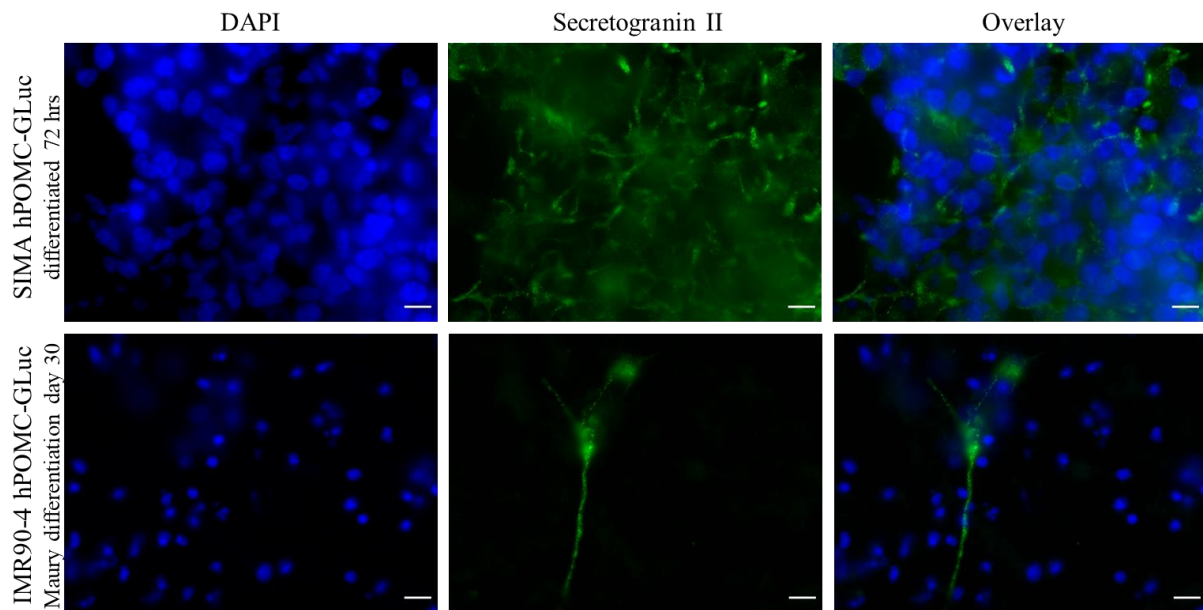
Figure 53: Differentiated SIMA and IMR90-4 hPOMC-GLuc (Maury Day30) clones stained with IF indirect labelling method. Cells were fixed, permeabilized and then incubated overnight with a 1:500 dilution of mouse-α-SgII primary antibody. Secondary antibody incubation with Alexa-Fluor α-mouse 488 diluted to 1:1500. Cells were additionally incubated with DAPI for nuclei staining. scale bar = 10μm

Synaptic vesicles

The staining of mouse-α-synaptophysin (Syp) at a 1:200 dilution, as recommended by the manufacturer, was similar to that of SgII. The antibody successfully stained in axonal outgrowths and in clusters similar to those of proteins sorted through the Golgi apparatus in differentiated SIMA hPOMC-GLuc cells (top panel). In CRISPR-modified motor neurons differentiated according to Maury *et al* for 30 days (bottom panel) the antibody staining displayed very distinct and intense clusters of the protein which appeared to be along the axonal outgrowth, as well as some staining closer to the soma (Figure 54).

| | DAPI | Synaptophysin | Overlay |
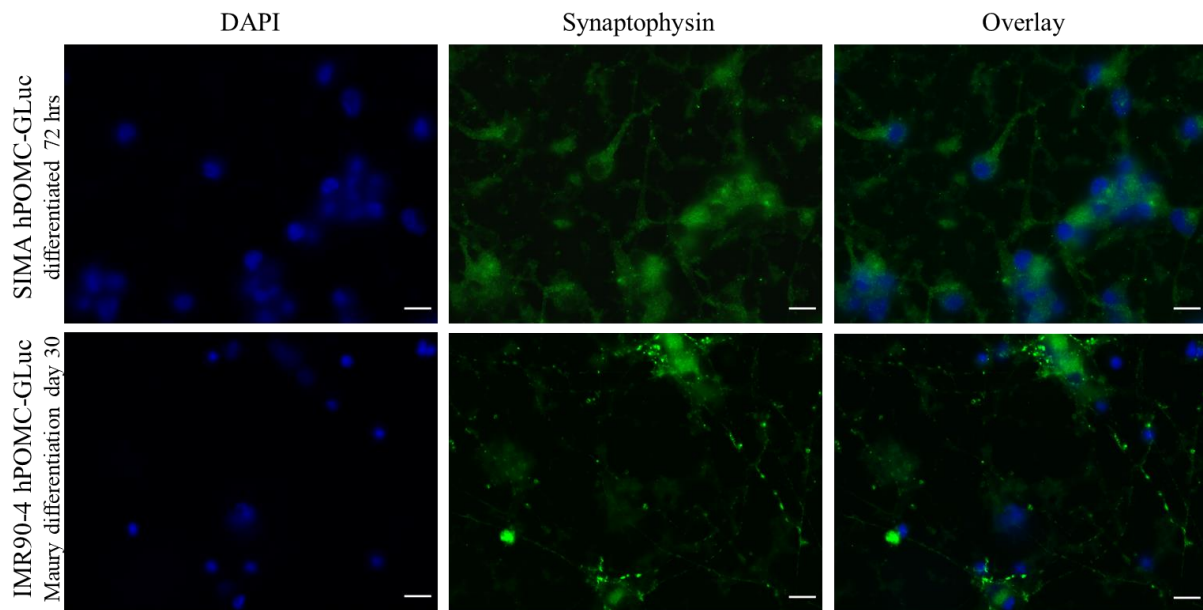
Figure 54: Differentiated SIMA and IMR90-4 hPOMC-GLuc (Maury Day30) clones stained with IF indirect labelling method. Cells were fixed, permeabilized and then incubated overnight with a 1:200 dilution of mouse-α-Syp primary antibody. Secondary antibody incubation with Alexa-Fluor α-mouse 488 diluted to 1:1500. Cells were additionally incubated with DAPI for nuclei staining. scale bar = 10µm

Cytosol

In order to establish an appropriate marker to visualized the cytosol, the antibody mouse-α-GAPDH was tested in two concentrations, both a 1:50 dilution (Appendix 6.6.3.2, Figure 147) and a 1:250 dilution (not shown), neither condition stained successfully and this antibody was not considered in any subsequent analyses.

Motor neurons

Islet1 is a transcription factor that is expressed in neuronal cells only after differentiation into motor neurons. The mouse-α-Islet1 clone [1B1] antibody test labelled all cells, including non-differentiated cells, non-specifically (see Appendix, 6.6.3.2). The mouse-α-Islet1 clone [1B1] antibody was not used further in this project. The mouse-α-Islet1 clone [1H9] antibody was tested at a dilution of 1:200 and appeared to react specifically. The antibody does not non-specifically label undifferentiated iPSCs (top panel Figure 55). One cell, indicated by an arrow, was labeled in the CRISPR-modified motor neurons differentiated according to Maury *et al* for 30 days (bottom panel Figure 55). Since only one cell was labeled with the mouse-α-Islet1 [1H9] antibody, apparently the differentiation protocol into motor neurons is not very efficient.
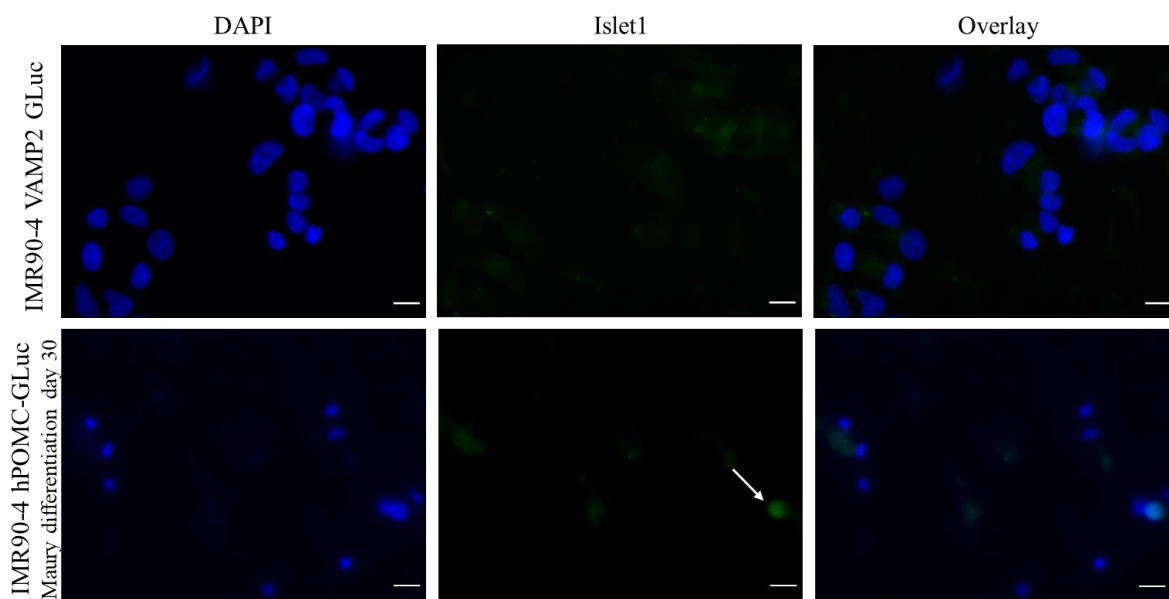
Figure 55: Undifferentiated IMR90-4 VAMP2-GLuc clone and differentiated IMR90-4 hPOMC-GLuc clone (Maury Day30) stained with IF indirect labeling method. Cells were fixed, permeabilized and then incubated overnight with a 1:200 dilution of mouse-α-Islet1 primary antibody [1H9]. Secondary antibody incubation with Alexa-Fluor α-mouse 488 diluted to 1:1500. Arrow points to Islet1 positive cell in differentiated IMR90-4 hPOMC-GLuc clone. Cells were additionally incubated with DAPI for nuclei staining. scale bar = 10µm

3.3 Production and validation of GLuc clones

3.4 Characterization of GLuc clones

Once the stable expression of GLuc in the clones was established, demonstrating the potential functionality of the clones in the MoN-Light BoNT luciferase release assay, it is necessary to completely characterize each clone using the methods described above to verify:

- The integration of the GLuc into the AAVS1 safe harbor locus

- The single integration of GLuc into the gDNA, without any off-target events

- The localization of the signal peptide tagged GLuc in secretory vesicles

3.4.1 Confirmation of integration at AAVS1 safe harbor locus

The integration of the donor DNA at the AAVS1 safe harbor locus was confirmed by PCR as described in the optimization section 3.2.1. The SIMA hPOMC-GLuc clones almost all contained at least one copy of the donor DNA at the AAVS1 safe harbor locus, indicated by the existence of a PCR product band at 3000 bp for the clones SIMA hPOMC-GLuc clones 2, 4, 5, 29, 32 (Figure 56, top gel) and SIMA hPOMC-GLuc clones 3, 4, 6, 11, 13, 18, 19, 21, 22, 23, 26, 27, and 28 (Figure 56, bottom gel). As expected, the amplification of both the non-transfected SIMA gDNA (top gel well 1) and the SIMA Random-Insertion_hPOMC1-26GLuc prototype DNA with random GLuc insertions (top gel well 8) only produced PCR products

visible at 850 bp, since both these cell lines have WT alleles at the AAVS1 safe harbor locus. Interestingly, almost all of the SIMA hPOMC-GLuc clones also contained a WT amplicon at 850 bp, except for clone 29 (well 6, top gel). These clones with two products are putative heterozygotes. Clone 29 (well 6, top gel) had a strong product at the insert location, and no product at the WT location, suggesting that this clone might contain a homozygous insertion of the donor DNA at the AAVS1 safe harbor locus. However there are several light bands visible between 850 bp and 3000 bp, which do not appear in the products of the remaining clones adding some uncertainty to these results. Furthermore, all the clones positive for the 3000 bp insert amplicon also contained a faint and unspecific amplicon at approximately 1600 bp (Figure 56). The expected product sizes of the WT allele amplicon and the insert amplicon are indicated by white arrows.
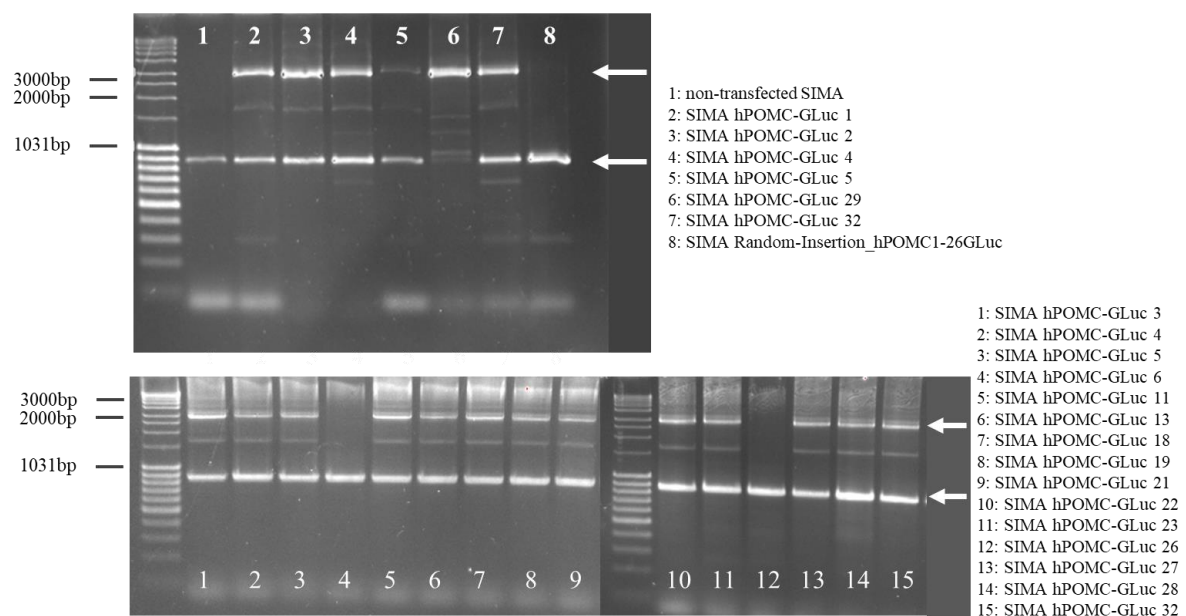
Figure 56: Donor DNA insert confirmation at AAVS1 safe harbor locus in SIMA hPOMC-GLuc clones and SIMA Random-Insertion_hPOMC1-26GLuc. Products amplified by PCR with the primer pair 5'probeWT_F and Rpcr-wt-3'HA with an annealing temperature of 70 °C. Expected product sizes WT: 850 bp, insert: 3000 bp.

Fewer than expected of the IMR90-4 hPOMC-GLuc clones contained donor DNA at the AAVS1 safe harbor locus. Clones 2, 3, and 7 all only contained WT alleles, with PCR products at 850 bp. The non-transfected IMR90-4 control also only produced an amplicon at 850 bp. Clone 1 appeared to be a heterozygote with amplicons at both the insert and WT product sizes. Both the clones 4 and 6 appeared to have homozygous insertions, with only one distinct band at 3000 bp. The PCR reaction for clone 5 did not result in any successful amplification (Figure 57). The expected product sizes of the WT allele amplicon and the insert amplicon are indicated by white arrows.

108

Section 3.4
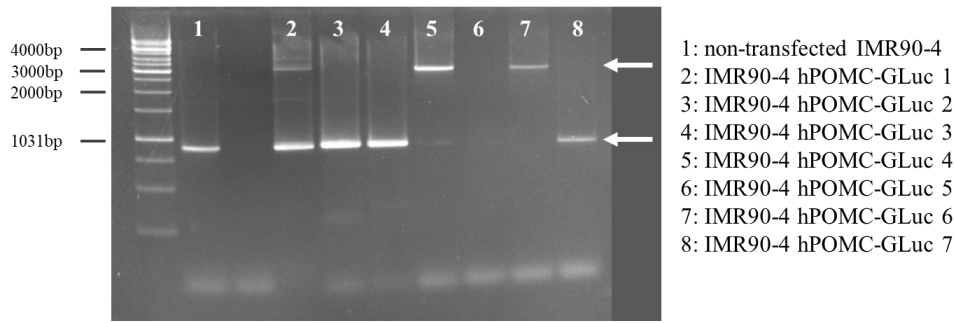
Figure 57: Donor DNA insert confirmation at AAVS1 safe harbor locus in IMR90-4 hPOMC-GLuc clones. Products amplified by PCR with the primer pair 5'probeWT_F and Rpcr-wt-3'HA with an annealing temperature of 70 °C. Expected product sizes WT: 850 bp, insert: 3000 bp.

In contrast to the IMR90-4 hPOMC-GLuc clones, which appeared to rarely incorporate the donor DNA at the correction insertion locus, all of the IMR90-4 no tag GLuc clones contained at least one allele with the correct insertion. The correct amplicon size for the no tag GLuc insert is 2925 bp. IMR90-4 no tag GLuc clones 1, 3, 4, 7, 9, 10, and 12 were all homozygote insert clones, with only an amplicon at 2925 bp and no amplified product from the WT allele. Clones 2, 5, 6, 9, and 11 all contained a band at the insert amplification size, as well as the WT amplification size, suggesting that these clones have a heterozygous insertion (Figure 58). The expected product sizes of the WT allele amplicon and the insert amplicon are indicated by white arrows.
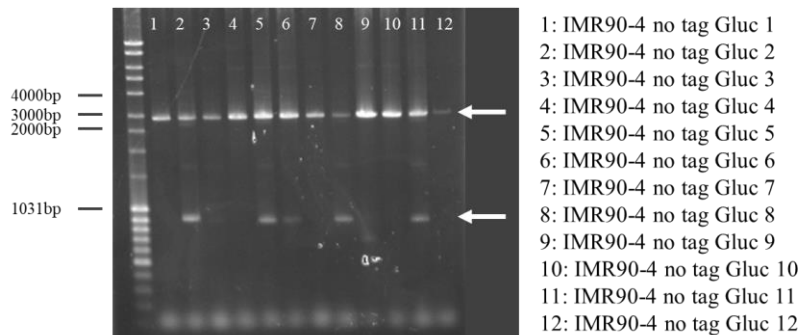
Figure 58: Donor DNA insert confirmation at AAVS1 safe harbor locus in IMR90-4 no tag GLuc clones. Products amplified by PCR with the primer pair 5'probeWT_F and Rpcr-wt-3'HA with an annealing temperature of 70 °C. Expected product sizes WT: 850 bp, insert:2925 bp.

Similar to the results of the IMR90-4 no tag GLuc clones, almost all the IMR90-4 CgA-GLuc clones contained an insert at the AAVS1 safe harbor locus. The correct insert confirmation size for the IMR90-4 CgA-GLuc clone is 4289 bp, due to the much larger CgA sorting sequence. The IMR90-4 CgA-GLuc clones 1, 3, 4, 6, 7, 14, and 15 all appeared to have homozygous insertions with only one amplicon at 4289 bp. The clones 11 and 16 both lacked an insertion at the AAVS1 safe harbor locus. The remaining clones all appeared to have heterozygous insertions of the donor DNA at the AAVS1 safe harbor locus, with amplicons both at the expected WT and insertion product sizes (Figure 59). The PCR products and the two

gels displayed in Figure 57 were amplified and separated by gel electrophoresis on two separate days, for this reason the exposure of the gel and separation of the products are different. The expected product sizes of the WT allele amplicon and the insert amplicon are indicated by white arrows.
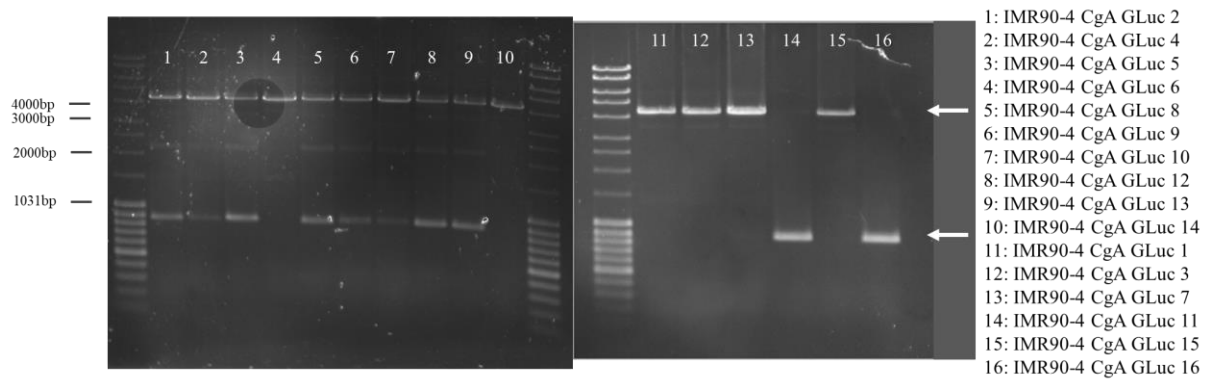
Figure 59: Donor DNA insert confirmation at AAVS1 safe harbor locus in IMR90-4 CgA-GLuc clones. Products amplified by PCR with the primer pair 5'probeWT_F and Rpcr-wt-3'HA with an annealing temperature of 67.5 °C. Expected product sizes WT: 850 bp, insert: 4289 bp.

The IMR90-4 SgII-GLuc clones almost all contained at least one copy of the insert at the AAVS1 safe harbor locus. For this particular donor DNA amplification was expected at 4757 bp. Only the IMR90-4 SgII-GLuc clone 14 contained an insertion band that was only at 2500 bp, indicating a problem with the integration of the donor DNA in this clone. The IMR90-4 SgII-GLuc clones 1, 2, 5, 7, and 8 appeared to have heterozygous insertions with amplicons at both 4757 bp and 850 bp, while the rest of the clones were putative homozygotes (Figure 60). The PCR products and the two gels displayed in Figure 58 were amplified and separated by gel electrophoresis on two separate days, for this reason the exposure of the gel and separation of the products are different. The expected product sizes of the WT allele amplicon and the insert amplicon are indicated by white arrows.
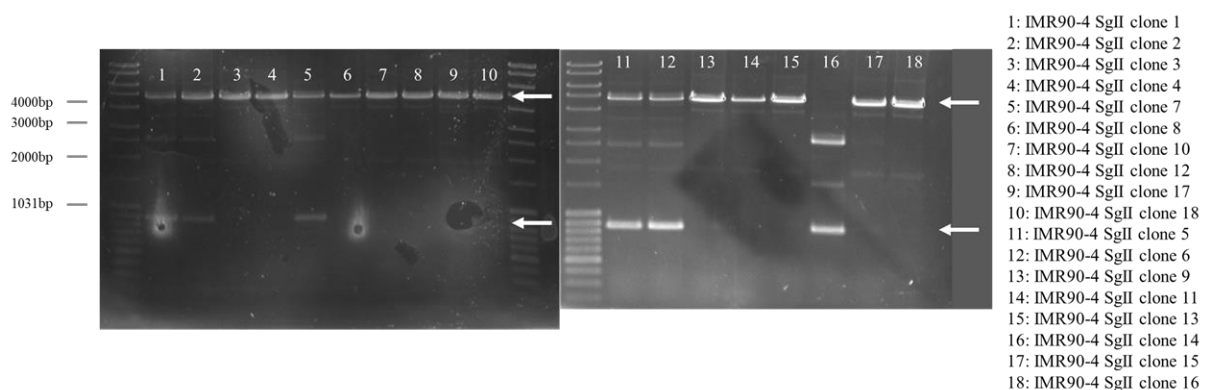
Figure 60: Donor DNA insert confirmation at AAVS1 safe harbor locus in IMR90-4 SgII-GLuc clones. Products amplified by PCR with the primer pair 5'probeWT_F and Rpcr-wt-3'HA with an annealing temperature of 67.5 °C. Expected product sizes WT: 850 bp, insert: 4757 bp.

110

The expected insert size for the SIMA VAMP2-GLuc clones was at 3292 bp. All clones contained amplified products at both the correct insertion size and the correct WT size, implicating that these clones had correct insertions of the donor DNA in only one allele, while the other allele was unaffected by the homologous recombination. Similar to the insert confirmation of the SIMA hPOMC-GLuc clones, there was an extra non-specific band in the amplification of the SIMA VAMP2-GLuc clones. This non-specific band was found at 2000 bp (Figure 61). The expected product sizes of the WT allele amplicon and the insert amplicon are indicated by white arrows.

1: SIMA VAMP2-GLuc 1
2: SIMA VAMP2-GLuc 4
3: SIMA VAMP2-GLuc 5
4: SIMA VAMP2-GLuc 6
5: SIMA VAMP2-GLuc 8
6: SIMA VAMP2-GLuc 9
7: SIMA VAMP2-GLuc 14
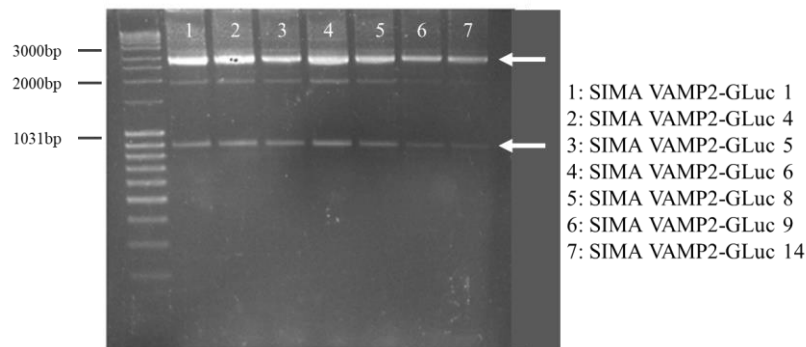
Figure 61: Donor DNA insert confirmation at AAVS1 safe harbor locus in SIMA VAMP2-GLuc clones. Products amplified by PCR with the primer pair 5'probeWT_F and Rpcr-wt-3'HA with an annealing temperature of 70 °C. Expected product sizes WT: 850 bp, insert: 3292 bp.

The expected insert size for the IMR90-4 VAMP2-GLuc clones was also at 3292 bp. The majority of the IMR90-4 VAMP2-GLuc clones were homozygous for the donor DNA insertion at the AAVS1 safe harbor locus. Only clone number 10 did not appear to carry any allele containing the donor DNA at the AAVS1 safe harbor locus, and clone 16 appeared to carry heterozygous alleles for both the insert and WT sequences (Figure 62). The expected product sizes of the WT allele amplicon and the insert amplicon are indicated by white arrows.
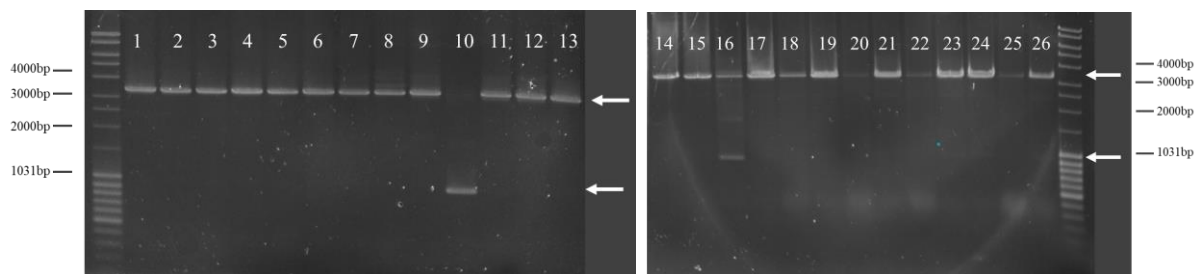
Figure 62: Donor DNA insert confirmation at AAVS1 safe harbor locus, VAMP2-GLuc clones 1 through 26. Products amplified by PCR with the primer pair 5'probeWT_F and Rpcr-wt-3'HA with an annealing temperature of 70 °C. Expected product sizes WT: 850 bp, insert: 3292 bp.

3.4.1.1 Sanger sequencing of homozygote clones

Sanger sequencing was performed with the insert confirmation PCR products of the two clone groups which would be used in the functional MoN-Light assay, IMR90-4 hPOMC-GLuc

and IMR90-4 VAMP2-GLuc. Because the product amplified for the insert confirmation test did not completely span the complete 5' of the left HA to 3' of the right HA range, two different amplicons were sequenced for each clone. The first amplicon was the exact product from the insert confirmation test. The second amplicon overlapped the first amplicon in the Ef1-HTLV promoter and then extended into genomic sequence 3' of the right homology arm (Figure 63).
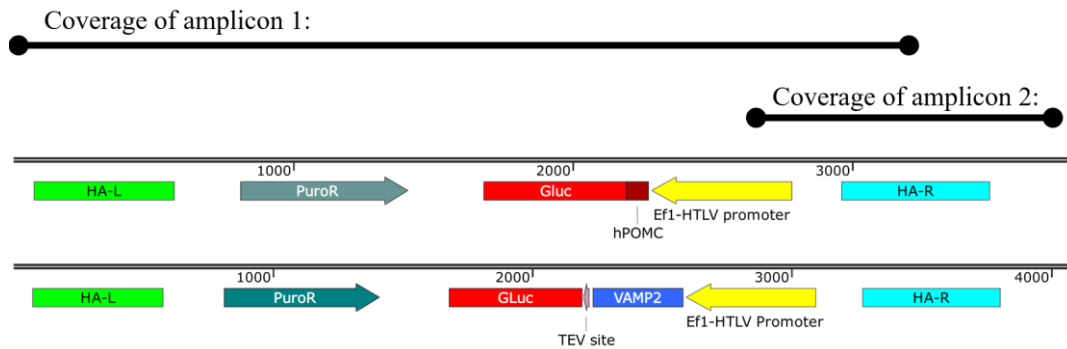
Figure 63: Summary of hPOMC-GLuc and VAMP2-GLuc donor DNA segments after integration into the cell's genomic DNA and amplicons used for Sanger sequencing.

Sequencing of IMR90-4 hPOMC-GLuc homozygote clones

Two IMR90-4 hPOMC-GLuc clones were identified as homozygotes by the insert confirmation PCR. Amplicons 1 and 2 (Figure 63) for both IMR90-4 hPOMC-GLuc clones 4 and 6 were Sanger sequenced. The readable range of sequencing for clone 4 was almost identical to that of clone 6 and is therefore not summarized here (see Appendix 6.4.11). In the summary of the sequence coverage of the homology arms and donor DNA insert for clone 6, the arrows filled in dark red represent the areas of unequivocal reads of the genomic DNA (Figure 64). All of the important regions, including the left and right homology arms, the Ef1-HTLV promoter, and the hPOMC-GLuc sequences were verified by at least one successful and unequivocal sequencing run. The only region which fell outside the sequencing range was a segment of the puromycin resistance gene and part of the gene's 3' flanking region. A C>T nucleotide change was identified at genomic position chr19:55115458 in the 3'HA, this base change was found in both clones 6 and 4. This region of the 3'HA in the donor DNA plasmids was not covered by Sanger sequencing, however the plasmid reference sequence indicates that the corresponding base in the 3'HA of the plasmid is a T. Since the non-transfected WT allele was not sequenced, it is unclear if the base change is inherent to the IMR90-4 cells or if it was incorporated into the gDNA from the plasmid template by homologous recombination. Other than this single base change, the remaining region of the right homology arm and the entire regions of the left homology arm, hPOMC-GLuc sequence, and the Ef1-HTLV promoter

112

conformed to the expected reference sequence of the genomic DNA and the integrated donor DNA. The sequencing confirmed that the donor DNA was successfully integrated without errors into both alleles of this homozygous clone (see Appendix, 6.4.10 for complete alignment data).
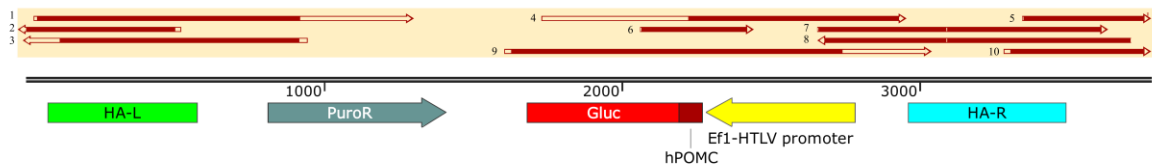
Figure 64: Sanger sequencing coverage of hPOMC-GLuc clone 6, donor DNA inserted into AAVS1 safe harbor locus.

Sequencing of IMR90-4 VAMP2-GLuc homozygote clone

One IMR90-4 VAMP2-GLuc clone, number 11 (Figure 62), was selected for sequence confirmation. The same amplicons were prepared as for the IMR90-4 hPOMC-GLuc sequencing, as described in Figure 63. The sequencing coverage is visualized by the dark red arrows at the top of the figure, indicating the area of the gDNA which was successfully and unequivocally sequenced (Figure 65). The coverage included the entire left and right homology arms, as well as all of the donor DNA, except for a short segment of the puromycin resistance gene. The only discrepancy of the sequenced sample from the reference sequence was the same C>T nucleotide change found in the sequences of both hPOMC-GLuc clones. The sequencing once again confirmed that the donor DNA was successfully integrated without errors into both alleles of this homozygous clone (see Appendix, 6.4.12 for complete alignment data).
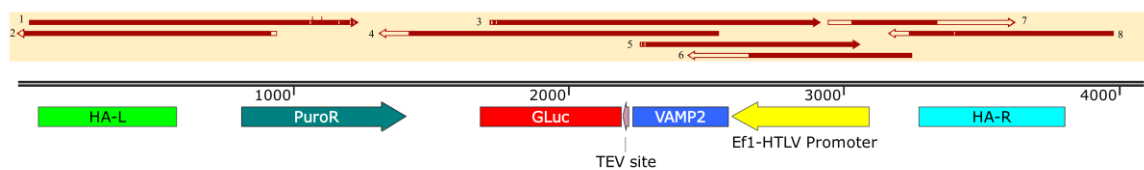
Figure 65: Sanger sequencing coverage of VAMP2-GLuc clone 11, donor DNA inserted into AAVS1 safe harbor locus.

3.4.1.2   Summary of insert confirmation at AAVS1 safe harbor locus

The results of the insertion of the donor DNA in the AAVS1 safe harbor locus for a selected clone representative for each transfection construct are summarized below. The donor DNA is depicted as a green segment in the DNA double helix. Both alleles are presented per clone. The IMR90-4 hPOMC-GLuc clone 6, IMR90-4 no tag GLuc clone 4, IMR90-4 SgII-GLuc clone 11, and IMR90-4 VAMP2-GLuc clone 11 all contained homozygous integrations of the donor DNA in the AAVS1 safe harbor locus. On the other hand, the SIMA hPOMC-

GLuc clones 1, 2, 4, and 5, IMR90-4 CgA-GLuc clone 8, and SIMA VAMP2-GLuc clone 1 all only incorporated the donor DNA into a single allele at the AAVS1 safe harbor locus (Figure 66). A selection of four SIMA hPOMC-GLuc clones were initially selected in order to verify the luciferase release assay in multiple clones.

	theoretical GLuc localization	insertion AAVS1 locus
SIMA hPOMC-GLuc (clones 1, 2, 4, 5)		allele 1 - contains donor DNA allele 2 - contains WT DNA
IMR90-4 hPOMC-GLuc (clone 6)		allele 1 - contains donor DNA allele 2 - contains donor DNA
IMR90-4 no tag GLuc (clone 4)		allele 1 - contains donor DNA allele 2 - contains donor DNA
IMR90-4 CgA-GLuc (clone 8)		allele 1 - contains donor DNA allele 2 - contains WT DNA
IMR90-4 SgII-GLuc (clone 11)		allele 1 - contains donor DNA allele 2 - contains donor DNA
SIMA VAMP2-GLuc (clone 1)		allele 1 - contains donor DNA allele 2 - contains WT DNA
IMR90-4 VAMP2-GLuc (clone 11)		allele 1 - contains donor DNA allele 2 - contains donor DNA

Figure 66: Summary of AAVS1 safe harbor locus insert confirmation for a selected CRISPR-modified clone for each construct. Donor DNA is depicted as a green segment in the DNA double helix.

3.4.2 Analysis of off-target donor DNA integrations

The second important factor in the characterization of the CRISPR/Cas9 manipulated cells, is to verify that there were no off-target integrations of the donor DNA in the cell's genome. The method, dc-qcnPCR, established to quantify the copy number of GLuc in genomic DNA, was fully explained in the optimization section (3.2.2.3).

While SIMA hPOMC-GLuc clones were all heterozygotes for the insertion of the donor DNA, it is clear from the copy number quantification that in some of the clones there was far more than one insert of donor DNA somewhere else in the genome. GLuc amplification compared to *RBBP7* in both clones 4 and 5 was n = 1.7 and n = 2, respectively, suggesting a single off-target integration event other than the heterozygous insertion of the donor DNA at the AAVS1 safe harbor locus. The remaining clones contained a mix of both single integrations and off-target events, especially exemplified by clone 29, which had an n-fold of *RBBP7* of more than 100. In this figure, the copy number of GLuc in the SIMA Random-Insertion_hPOMC1-26GLuc prototype was also assessed. Unsurprisingly, there were several random insertions of GLuc in the genome of these cells, with an n-fold of *RBBP7* of 4 (Figure 67 A). The ratio of the copy number of *RBBP7/CHOP* for control females was on average n = 1.1 and for control males was on average 0.57, verifying the efficacy of this experiment: the males have half the number of copies of the ChrX gene compared to the females (Figure 67 B).
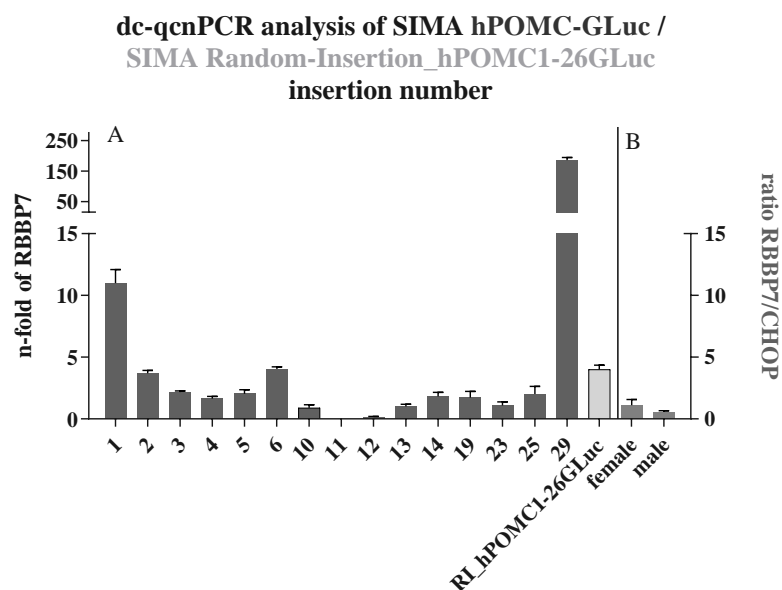
Figure 67: Copy number analysis of GLuc in SIMA hPOMC-GLuc clones and SIMA Random-Insertion_hPOMC1-26GLuc (RI_hPOMC1-26GLuc). A: copy number calculated by ΔCt of GLuc normalized to corresponding plasmid amplification, CHOP and ChrX amplification. B: ratio of the copy number calculated by ΔCt of RBBP7, normalized to CHOP in female and male control gDNA. Values displayed mean ± SD, n = 3.

The number off-target integrations of GLuc in the IMR90-4 hPOMC-GLuc clones was relatively lower than the off-target integrations in the SIMA clones. In fact, the assay indicated that clone 1 only had one copy of GLuc integrated into its genome, since it contained half of the number of copies of GLuc than it did the number of copies of *RBBP7* (n = 2 in the female-derived IMR90-4 cells). Clone 3 appeared to have no integrations of GLuc in the genome, while both clones 4 and 6 had the same number of copies of GLuc as they do *RBBP7*, therefore n = 2

(Figure 68 A). These numbers perfectly complimented the PCR insertion confirmation results, in which clone 1 only had a heterozygous insertion of the donor DNA and both clone 4 and clone 6 had homozygous insertions. Furthermore, the amplification of clone 3 had only resulted in amplification of WT alleles (Figure 57). This indicated that the clones 1, 4, and 6 all only contained on-target insertions of the donor DNA into the AAVS1 safe harbor locus, while clone 3 had no integration of the donor DNA into its gDNA. The ratio of the copy number of *RBBP7/CHOP* for control females was on average n = 1.0 and for control males was on average 0.55, verifying the efficacy of this experiment: the males have half the number of copies of the ChrX gene compared to the females (Figure 68 B).
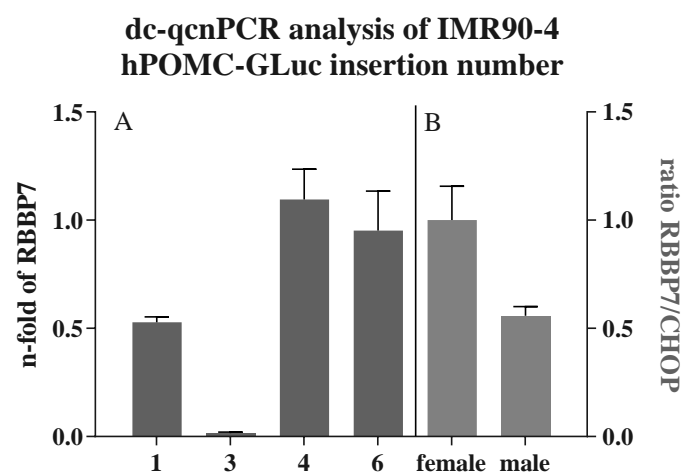
Figure 68: Copy number analysis of GLuc in IMR90-4 hPOMC-GLuc clones. A: copy number calculated by ΔCt of GLuc normalized to corresponding plasmid amplification, CHOP and ChrX amplification. B: ratio of the copy number calculated by ΔCt of RBBP7, normalized to CHOP in female and male control gDNA. Values displayed mean ± SD, n = 3.

The copy number of GLuc in the IMR90-4 no tag GLuc clones was also consistently low, however, it appeared that a few of the clones contained probable off-target integrations. The copy number assay indicated that, for example, IMR90-4 no tag GLuc clone 2 contained two copies of GLuc in its gDNA (Figure 69 A), however, as seen in the insert confirmation gel (Figure 58), this clone was only a heterozygote for the donor DNA insertion at the AAVS1 safe harbor locus. Therefore one copy of the donor DNA was definitely found at the AAVS1 safe harbor locus, but there was likely another copy that had been integrated off-target in the genome. All the other clones reached at least 1.5 fold of the *RBBP7* copy number, indicating that they too likely had 1 to 2 off-target integrations in their genomes (Figure 69 A). None of the selected IMR90-4 no tag GLuc clones contained only on-target donor DNA integrations. The ratio of the copy number of *RBBP7/CHOP* for control females was on average n = 1.1 and

for control males was on average 0.6, verifying the efficacy of this experiment: the males have half the number of copies of the ChrX gene compared to the females (Figure 69 B).
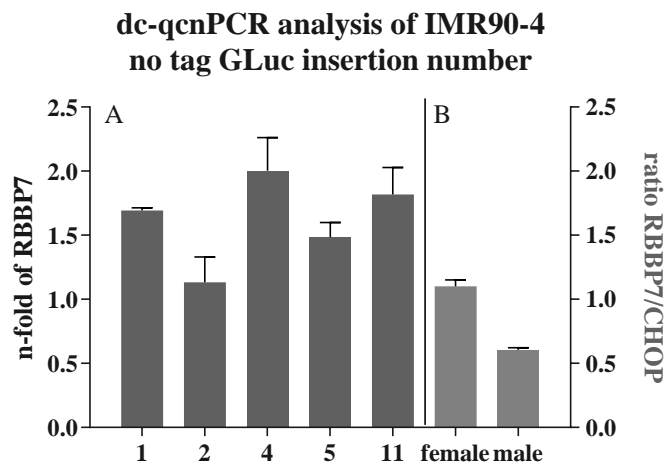
Figure 69: Copy number analysis of GLuc in IMR90-4 no tag GLuc clones. A: copy number calculated by ΔCt of GLuc normalized to corresponding plasmid amplification, CHOP and ChrX amplification. B: ratio of the copy number calculated by ΔCt of RBBP7, normalized to CHOP in female and male control gDNA. Values displayed mean ± SD, n = 3.

The low number of integrations detected in the IMR90-4 clones remained consistent after consideration of the IMR90-4 CgA-GLuc clones. The clones 11 and 16 did not appear to have incorporated any copies of GLuc, neither on-target nor off-target (Figure 70 A), data which was supported by the amplification of only WT alleles in the insert confirmation PCR (Figure 59). The clones 8 and 13 both had the same copy number of GLuc as they do *RBBP7*, indicating two GLuc integration events. However, as seen in Figure 59, both these clones were identified as heterozygotes for the donor DNA insertion, which once again suggested an off-target integration of at least one copy. All the other clones had at least 1.5 fold copies of GLuc as compared to *RBBP7*, indicating that all clones of the CgA-GLuc series had at least one off-target donor DNA integration (Figure 70 A). The ratio of the copy number of *RBBP7/CHOP* for control females was on average n = 1.2 and for control males was on average 0.57, verifying the efficacy of this experiment: the males have half the number of copies of the ChrX gene compared to the females (Figure 70 B).
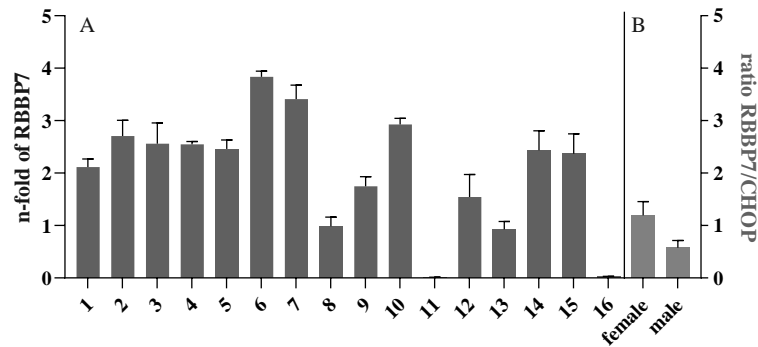
Figure 70: Copy number analysis of GLuc in IMR90-4 CgA-GLuc clones. A: copy number calculated by ΔCt of GLuc normalized to corresponding plasmid amplification, CHOP and ChrX amplification. B: ratio of the copy number calculated by ΔCt of RBBP7, normalized to CHOP in female and male control gDNA. Values displayed mean ± SD, n = 3.

There were several clones from the IMR90-4 SgII-GLuc clone group that had from 0.5 to approximately 1 fold copies of GLuc compared to *RBBP7*, indicating a relatively low donor DNA insertion rate in these clones (Figure 71). The clone 1 had half the number of copies of GLuc as compared to *RBBP7*. The insertion confirmation PCR identified the clone as a heterozygote for GLuc (Figure 60), making it likely that this clone only had one copy of the donor DNA in one allele at the AAVS1 safe harbor locus. Furthermore, the clones 3, 4, 11 and 14 all had from 0.74 to 1.15 average fold copies of GLuc compared to *RBBP7*. All of these clones, except for 14, were identified as homozygote insertion clones by the confirmation PCR suggesting a likely insertion of two copies of the donor DNA into each allele at the AAVS1 safe harbor locus. The clone 14 had unidentifiable products in the insertion confirmation gel, so would be excluded from future use, despite having a low insertion copy number. The remaining clones had up to approximately 2.5 fold copies of GLuc compared to *RBBP7*, all of these clones likely had a few off-target integrations in their genomes (Figure 71 A). The ratio of the copy number of *RBBP7/CHOP* for control females was on average n = 1.2 and for control males was on average 0.57, verifying the efficacy of this experiment: the males have half the number of copies of the ChrX gene compared to the females (Figure 71 B).

**dc-qcnPCR analysis of IMR90-4
SgII-GLuc insertion number**
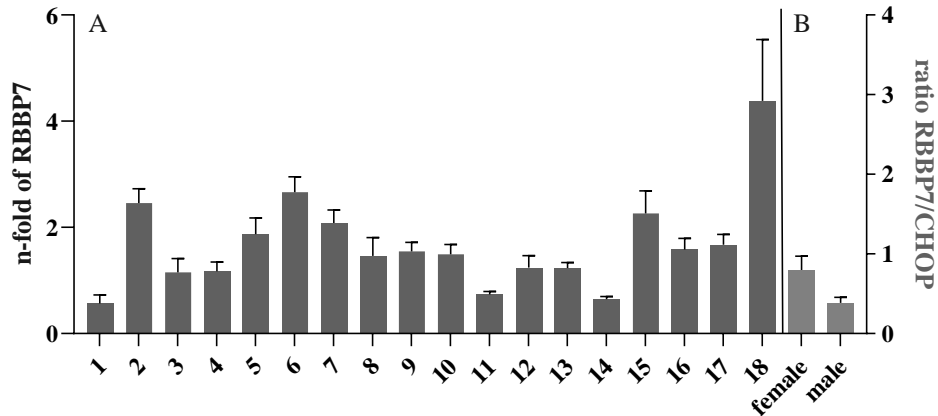
Figure 71: Copy number analysis of GLuc in IMR90-4 SgII-GLuc clones. A: copy number calculated by ΔCt of GLuc normalized to corresponding plasmid amplification, CHOP and ChrX amplification. B: ratio of the copy number calculated by ΔCt of RBBP7, normalized to CHOP in female and male control gDNA. Values displayed mean ± SD, n = 3.

Similar to the excessive donor DNA insertion rate found in some of the SIMA hPOMC-GLuc clones, the SIMA VAMP2-GLuc clones had between 16 to 50 fold copies of GLuc compared to *RBBP7*. All of these clones had a high number of off-target donor DNA insertions (Figure 72 A). The ratio of the copy number of *RBBP7/CHOP* for control females was on average n = 1.1 and for control males was on average 0.49, verifying the efficacy of this experiment: the males have half the number of copies of the ChrX gene compared to the females (Figure 72 B).

**dc-qcnPCR analysis of SIMA
VAMP2-GLuc insertion number**
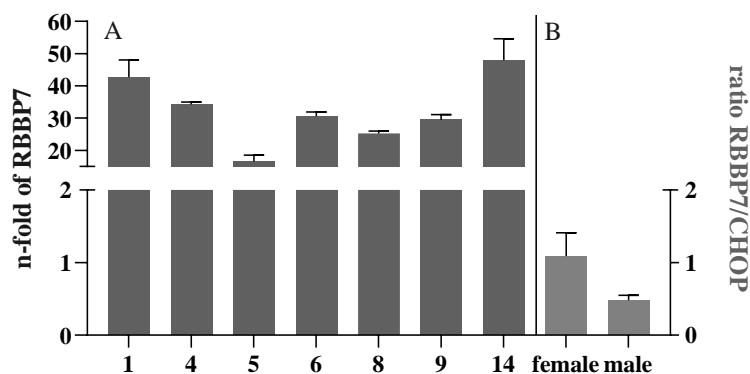
Figure 72: Copy number analysis of GLuc in SIMA VAMP2-GLuc clones. A: copy number calculated by ΔCt of GLuc normalized to corresponding plasmid amplification, CHOP and ChrX amplification. B: ratio of the copy number calculated by ΔCt of RBBP7, normalized to CHOP in female and male control gDNA. Values displayed mean ± SD, n = 3.

Once again, the number of off-target insertions detected in the iPSC line was much lower in comparison to the SIMA cell line. Almost all of the IMR90-4 VAMP2-GLuc clones had the same copy number of GLuc as compared to *RBBP7* (Figure 73 A). Interestingly, the

IMR90-4 VAMP2-GLuc clone 16, with only 0.5 fold copies of GLuc compared to *RBBP7* was also identified as a heterozygote donor DNA clone by the insert confirmation PCR (Figure 62 A) and therefore probably truly only had one copy of GLuc in the genome. IMR90-4 VAMP2-GLuc clone 10 had no identifiable copies of GLuc in this test, and also had homozygous WT alleles in the insert confirmation. Clones 2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 15, 17, 19, 20, and 22 were all identified as homozygotes for the donor DNA and had approximately equal copies of GLuc as compared to *RBBP7*, suggesting that these clones did not have any off-target GLuc integrations (Figure 73 A). The ratio of the copy number of *RBBP7/CHOP* for control females was on average n = 1.1 and for control males was on average 0.46, verifying the efficacy of this experiment: the males have half the number of copies of the ChrX gene compared to the females (Figure 73 B).

**dc-qcnPCR analysis of IMR90-4
VAMP2-GLuc insertion number**
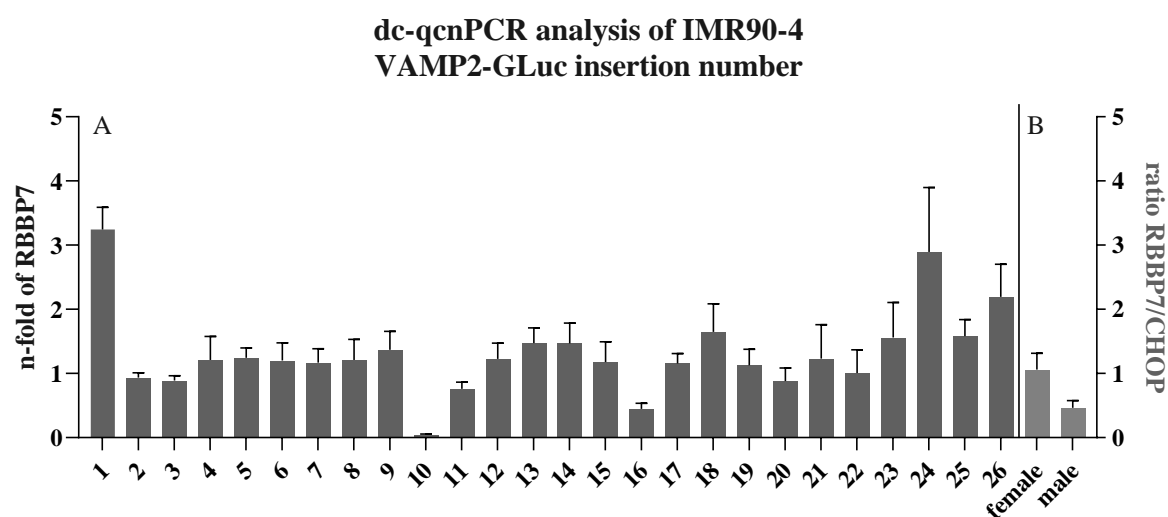
Figure 73: Copy number analysis of GLuc in IMR90-4 VAMP2-GLuc clones. A: copy number calculated by ΔCt of GLuc normalized to corresponding plasmid amplification, CHOP and ChrX amplification. B: ratio of the copy number calculated by ΔCt of RBBP7, normalized to CHOP in female and male control gDNA. Values displayed mean ± SD, n = 3.

3.4.2.1 Summary of off-target donor DNA integrations analyzed by dc-qcnPCR

The results in the identification of off-target donor DNA integrations for the selected clones representing each transfected construct is summarized below. In these particular clones, the IMR90-4 hPOMC-GLuc, SgII-GLuc, and VAMP2-GLuc clones did not contain any off-target donor DNA insertions (visualized by unaffected exemplary chromosomes). On the other hand, the SIMA hPOMC-GLuc, IMR90-4 no tag GLuc, CgA-GLuc, and the SIMA VAMP2-GLuc clones all contained integrations of the donor DNA in locations other than the AAVS1 safe harbor locus (visualized by exemplary chromosomes stamped with a green star, Figure 74). The IMR90-4 no tag GLuc clone 4 was chosen, despite having off-target donor DNA insertions,

120

because it was a homozygote clone, which propagated at a rate similar to the non-transfected IMR90-4 cells and showed no unusual spontaneous differentiation under standard cell culture care. Therefore it was considered to be an acceptable negative control.
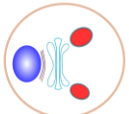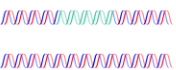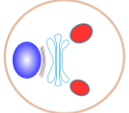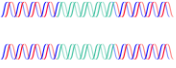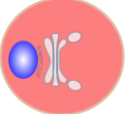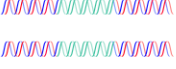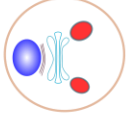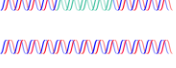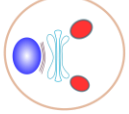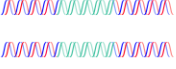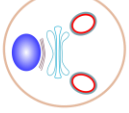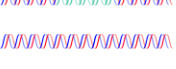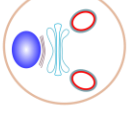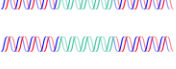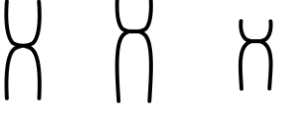
	theoretical GLuc localization	insertion AAVS1 locus	off-target donor DNA insertions?	
SIMA hPOMC-GLuc (clones 1, 2, 4, 5)				Yes
IMR90-4 hPOMC-GLuc (clone 6)				No
IMR90-4 no tag GLuc (clone 4)				Yes
IMR90-4 CgA-GLuc (clone 8)				Yes
IMR90-4 SgII-GLuc (clone 11)				No
SIMA VAMP2-GLuc (clone 1)				Yes
IMR90-4 VAMP2-GLuc (clone 11)				No

Figure 74: Continuation of summary of donor DNA integration results of selected CRISPR-modified clones representing each construct. Off-target insertion represented by presence or absence of green star on three exemplary chromosomes.

3.4.3 Localization of the signal peptide tagged GLuc by immunofluorescence

The third important step in the characterization of the clones developed for the MoN-Light assay is the confirmation that the GLuc protein is sorted into the secretory vesicles. The localization of GLuc was documented and compared with that of the Golgi apparatus and other naturally occurring proteins known to be sorted through the Golgi apparatus and into secretory vesicles. These results were summarized in a graph measuring the fluorescence intensities associated with the localization of each protein through the Z-axis of the cell. Furthermore, each

selected region was analyzed by the Fiji program JACoP, running a Pearson's colocalization analysis, corrected for fluorescent activity by the overlap coefficient, k1& k2, for each pixel[93,106]. These colocalization data were summarized at the end of this section.

SIMA hPOMC-GLuc colocalization

In order to analyze the localization of GLuc in the SIMA hPOMC-GLuc clone, the reporter protein GLuc (labeled in red) and CgA (labeled in green), a protein found in LDCVs, were identified by immunohistochemistry (Figure 75 A). The arrows in the overlay image indicate the regions analyzed for colocalization. Where the two proteins colocalized the color of the mixed channels became yellow/orange. The graphs tracked the intensity of fluorescence corresponding to GLuc and CgA along the Z-axis of the three randomly selected image sections (Figure 75 B). The fluorescence intensity and distribution through the cell were almost completely identical for the two proteins, with an average overlap correlation coefficient of 0.95 (Figure 75 C). Therefore the hPOMC-GLuc construct appeared to be successfully and efficiently sorting into LDCVs.
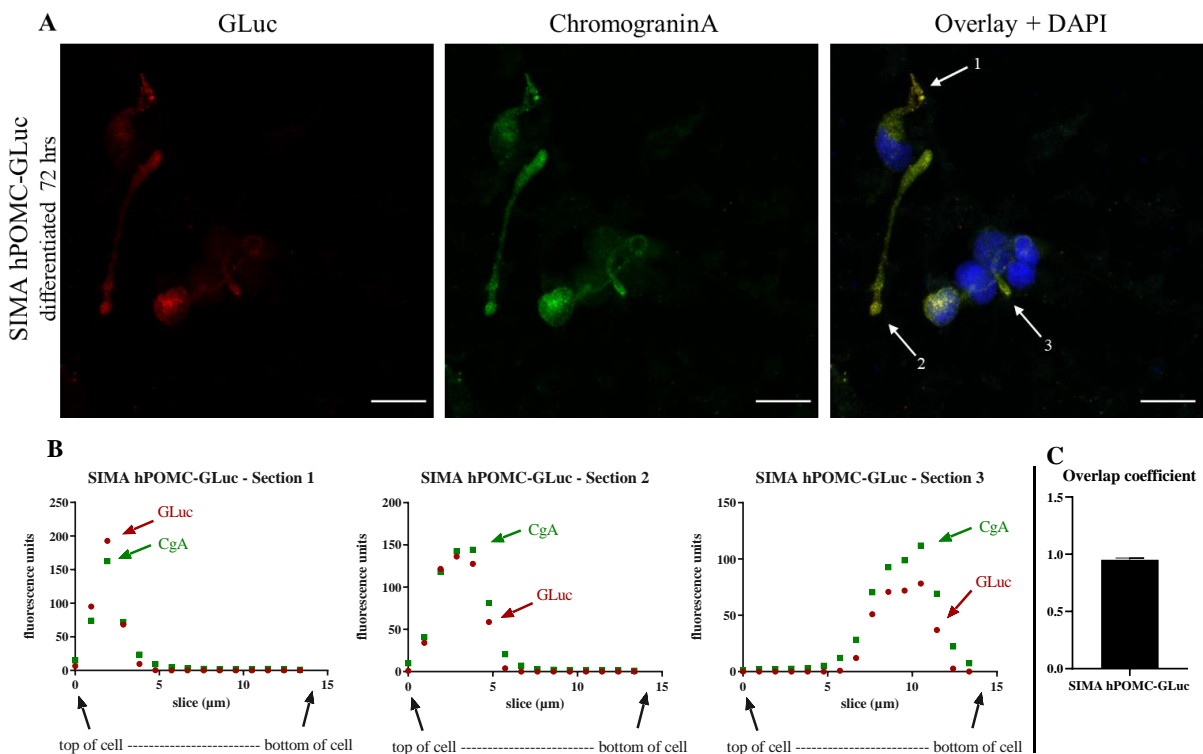
Figure 75: Colocalization of GLuc with CgA in SIMA hPOMC-GLuc clone. A: immunofluorescence double labeling of SIMA hPOMC-GLuc clone 5. Cells were fixed, permeabilized and then incubated overnight with rabbit-α-GLuc and mouse-α-CgA primary antibodies. Secondary antibody incubation with Alexa-fluor α-rabbit 568 (1:1000) and Alexa-fluor α-mouse 488 (1:1500). The overlay of all images includes DAPI labeled nuclei and sections of colocalization analysis indicated by white arrows in overlay image; scale bar = 20μm. Bottom panel: (B) graphical representation of colocalization analysis of GLuc and CgA by measurement of fluorescence units in selected cell sections of SIMA hPOMC-GLuc clone 5. (C) Overlap correlation coeffient of GLuc and CgA colocalization.

122

IMR90-4 hPOMC-GLuc colocalization

In order to analyze the cellular localization of GLuc in the IMR90-4 hPOMC-GLuc 6 clone, the proteins GLuc (labeled in red) and GM130 (labeled in green) were identified by immunohistochemistry (Figure 76 A). In this case, the iPSCs did not yet express CgA nor SgII, the two available antibodies that could be used to detect LDCVs. For this reason the GM130 antibody, which can be used to detect the *cis*-Golgi, was used to identify proteins that are moving through the secretory pathway towards vesicle packaging. The overlay image included the DAPI-stained nuclei, the arrows in the image indicate the regions analyzed for colocalization (Figure 76 A). The fluorescence intensity and distribution through the cell, visualized in the diagrams in the lower panel B, also illustrated an extremely similar pattern between GLuc and GM130 localization, with a peak of fluorescence approximately 7 µm thick through the cell. The average overlap coefficient between the fluorescence levels associated with these two proteins was 0.85 (Figure 76 C). The colocalization was slightly lower than that in the SIMA hPOMC-GLuc constructs associated with CgA, however it was still high enough to indicate that GLuc is sorted through the Golgi apparatus in preparation for sorting to LDCVs.
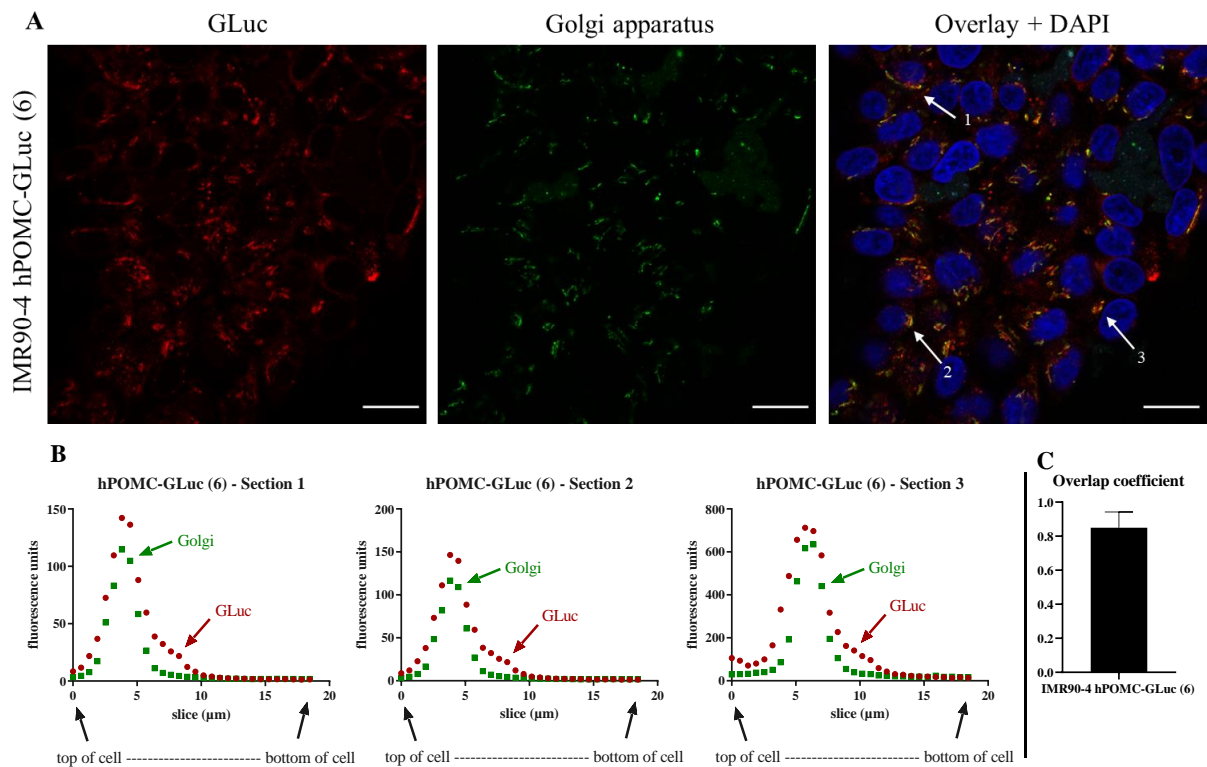
Figure 76: Colocalization of GLuc with Golgi in IMR90-4 hPOMC-GLuc clone 6. Panel A: immunofluorescence double labeling of IMR90-4 hPOMC-GLuc clone 6. Cells were fixed, permeabilized and then incubated overnight with rabbit-α-GLuc and mouse-α-GM130 primary antibodies. Secondary antibody incubation with Alexa-fluor α-rabbit 568 (1:1000) and Alexa-fluor α-mouse 488 (1:1500). The overlay of all images includes DAPI labeled nuclei and sections of colocalization analysis indicated by white arrows in overlay image; scale bar = 20µm. Lower panel: (B) Graphical representation of colocalization analysis of GLuc with GM130 by measurement of fluorescence units in selected cell sections in IMR90-4 hPOMC-GLuc (6). (C) Overlap coeffient of GLuc and GM130 colocalization.

Section 3.4

In order to validate that the high level of correlation between the localization of the GLuc and Golgi apparatus was not just by chance, the same experiment was carried out in a separate clone, the second homozygote IMR90-4 hPOMC-GLuc clone 4. Once again, the proteins GLuc (red) and GM130 (green) were identified by immunofluorescence. The overlay of these channels include an image of the DAPI-stained nuclei and the arrows indicating the regions of the image which were analyzed in detail for colocalization (Figure 77 A). The diagrams tracking the fluorescence intensity through the cell indicated that GLuc was partially found in the cytosol with a minor increase of fluorescence associated with GLuc outside of the fluorescence peak associated with the Golgi (Figure 77 B). Despite this, the majority of GLuc was colocalized with the Golgi apparatus, with strong peaks of protein localization at the exact same points in the cell (between 10 – 15 µm in each section). The average overlap correlation factor for the three sections was 0.79 (Figure 77 C), the small reduction in the correlation value was likely due to the small amount of cytosolic protein. Despite the reduction in correlation, the colocalization of GLuc and Golgi apparatus in a separate IMR90-4 hPOMC-GLuc clone also showed that this construct was sorted through the Golgi, destined for the LDCVs.
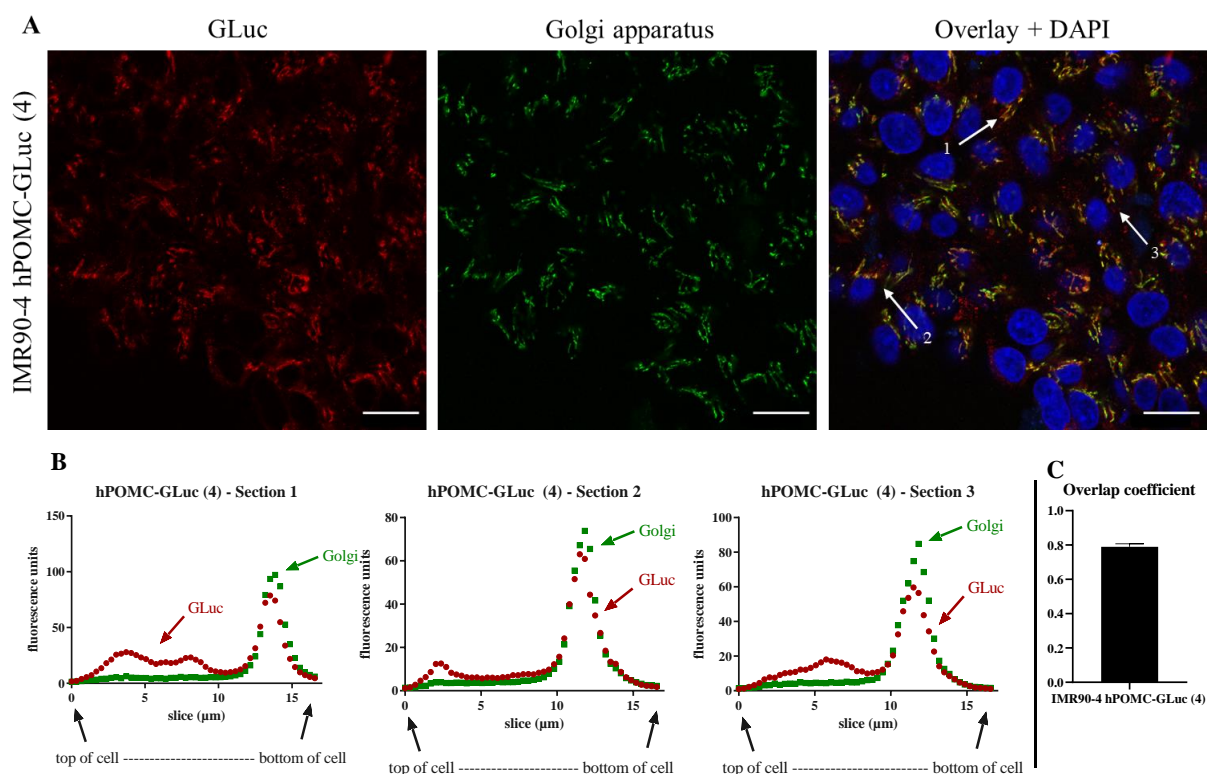
Figure 77: Colocalization of GLuc with Golgi in IMR90-4 hPOMC-GLuc clone 4. Panel A: immunofluorescence double labeling of IMR90-4 hPOMC-GLuc clone 4. Cells were fixed, permeabilized and then incubated overnight with rabbit-α-GLuc and mouse-α-GM130 primary antibodies. Secondary antibody incubation with Alexa-fluor α-rabbit 568 (1:1000) and Alexa-fluor α-mouse 488 (1:1500). The overlay of all images includes DAPI labeled nuclei and sections of colocalization analysis indicated by white arrows in overlay image; scale bar = 20µm. Lower panel: (B) Graphical representation of colocalization

124

Section 3.4

IMR90-4 no tag GLuc colocalization

The IMR90-4 no tag GLuc clones were created in order to quantify the amount of GLuc without a sorting signal that might colocalize by chance in the Golgi apparatus, in contrast to the signal peptide GLuc, which should be sorted through the Golgi apparatus. The distribution of GLuc (red) in no tag GLuc clone 4 was strikingly different from the GLuc distribution in the hPOMC-GLuc clones (Figure 78 A). There was no congregation of protein in specific areas of the cell, rather GLuc was distributed rather evenly throughout the cytosol. In the overlay of GLuc, Golgi (green), and the DAPI-stained nuclei, the arrows indicate the areas in which the colocalization of GLuc and Golgi was analyzed. The diagrams in the lower panel traced the fluorescence localization through the cell. The marker for the Golgi apparatus always spiked in one particular section of the cell. The fluorescence traces revealed that GLuc was partially colocalized with the spike in Golgi-associated fluorescence, however, it was found in greater quantities throughout the cell, independent of the Golgi apparatus (Figure 78 B). The average overlap correlation score was 0.55 (Figure 78 C). The low correlation coefficient confirmed the random distribution of GLuc when it is expressed in the no tag GLuc clone 4.
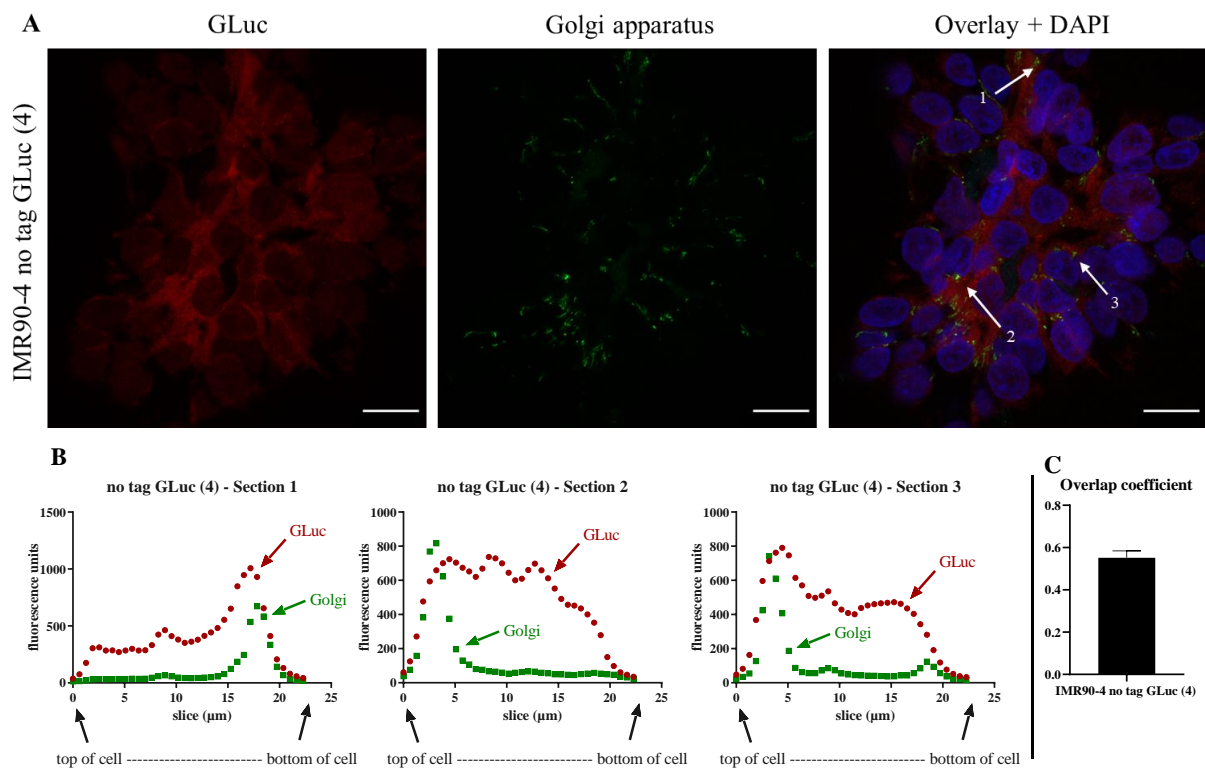
Figure 78: Colocalization of GLuc with Golgi in IMR90-4 no tag Gluc clone 4. Upper panel A: immunofluorescence double labeling of IMR90-4 no tag GLuc clone 4. Cells were fixed, permeabilized and then incubated overnight with rabbit-α-GLuc and mouse-α-GM130 primary antibodies. Secondary antibody incubation with Alexa-fluor α-rabbit 568 (1:1000) and Alexa-fluor α-mouse 488 (1:1500). The overlay of all images includes DAPI labeled nuclei and sections of colocalization analysis indicated by white arrows in overlay image; scale bar = 20μm. Lower panel: (B) Graphical representation of colocalization

125

Section 3.4

In order to validate that the low level of correlation between the localization of the GLuc and Golgi in one no tag GLuc clone was not just by chance, the same experiment was carried out in another separate clone, the IMR90-4 no tag GLuc clone 1. Once again, in this independent clone, the general appearance of GLuc distribution was very diffuse throughout the entire cell while the Golgi apparatus retained its strong perinuclear staining (Figure 79 A). In the overlay of GLuc (labeled in red) and Golgi (labeled in green), with the DAPI-stained nuclei, the arrows indicate the areas in which the colocalization of GLuc and Golgi was analyzed. The fluorescence traces showed that in sections 1 and 2, GLuc was found both colocalized with the typical Golgi spike and independent of the Golgi in the surrounding region. More drastically, in section 3 low levels of labeled GLuc appeared on either side of the Golgi spike, while the fluorescence levels decreased as soon as the Golgi labeling began (Figure 79 B). The average overlap coefficient for these sections was 0.45 (Figure 79 C). It was confirmed with this second no tag clone that the random distribution of GLuc had a much lower correlation with the Golgi apparatus than the hPOMC-GLuc clones. However, against expectations, no tag GLuc was not completely excluded from the Golgi apparatus.
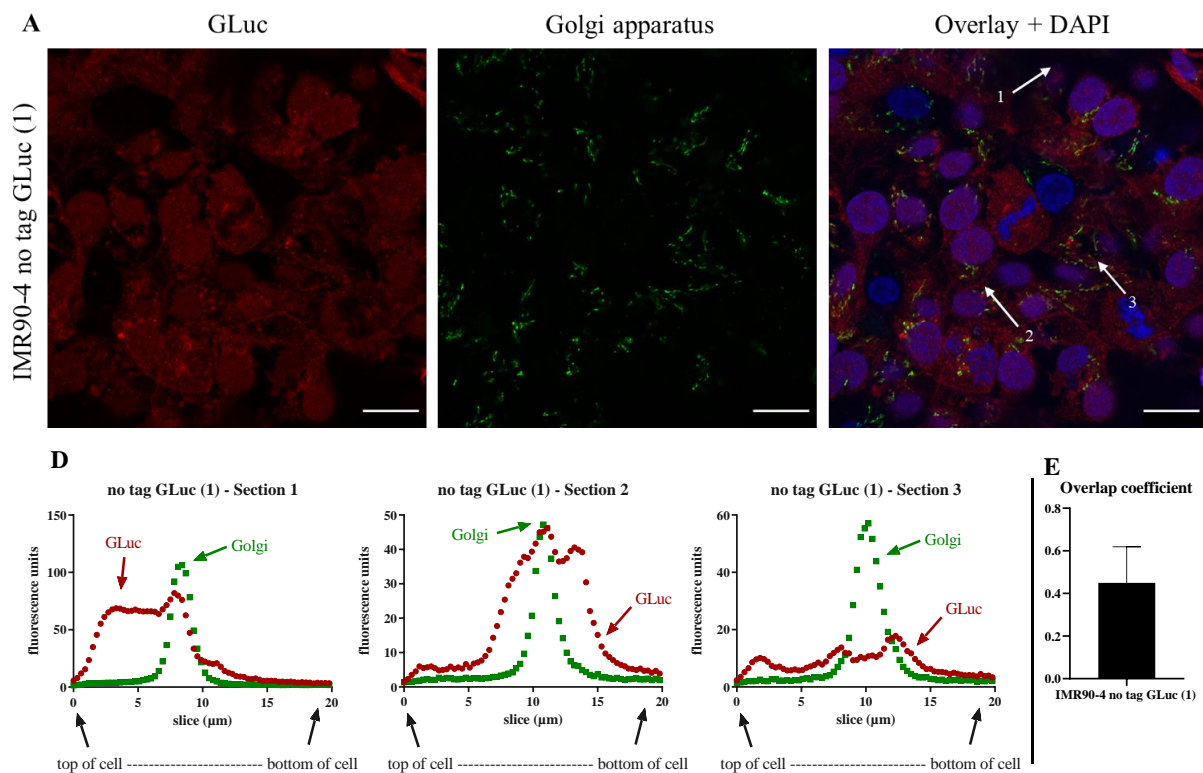
Figure 79: Colocalization of GLuc with Golgi in IMR90-4 no tag Gluc clone 1. Upper panel A: immunofluorescence double labeling of IMR90-4 no tag GLuc clone 1. Cells were fixed, permeabilized and then incubated overnight with rabbit-α-GLuc

and mouse-α-GM130 primary antibodies. Secondary antibody incubation with Alexa-fluor α-rabbit 568 (1:1000) and Alexa-fluor α-mouse 488 (1:1500). The overlay of all images includes DAPI labeled nuclei and sections of colocalization analysis indicated by white arrows in overlay image; scale bar = 20µm. Lower panel: (B) Graphical representation of colocalization analysis of GLuc with GM130 by measurement of fluorescence units in selected cell sections in IMR90-4 no tag GLuc (1). (C) Overlap coeffient of GLuc and GM130 colocalization.

IMR90-4 CgA-GLuc colocalization

The GLuc and GM130 proteins were also co-stained in CgA-GLuc clone 8. However, the GLuc staining was very fragmented and in some locations it was diffuse. The aggregation of the protein did not appear similar to the accumulation typical of the Golgi apparatus around perinuclear centrosomal region (Figure 80).
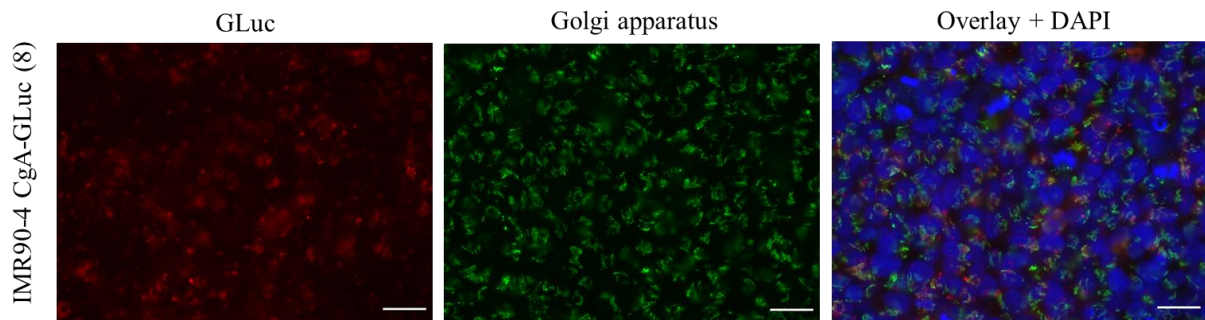
Figure 80: Immunofluorescence of GLuc and GM130 in IMR90-4 CgA clone. A: immunofluorescence double labeling of IMR90-4 CgA-GLuc clone 8. Cells were fixed, permeabilized and then incubated overnight with rabbit-α-GLuc and mouse-α-CgA primary antibodies. Secondary antibody incubation with Alexa-fluor α-rabbit 568 (1:1000) and Alexa-fluor α-mouse 488 (1:1500). The overlay of all images includes DAPI labeled nuclei. scale bar = 20µm.

The disruption in the CgA-GLuc protein distribution was compared directly with high magnification to the hPOMC-GLuc construct, since the CgA tagged clone was designed to sort GLuc to LDCVs analogous to hPOMC-GLuc. As described above, the expression of GLuc in the CgA clone aggregated in circular clusters surrounded by patches of diffuse protein. On the other hand, the GLuc distribution in the hPOMC-GLuc clone appeared in elongated clusters surrounding the nucleus (Figure 81). The CgA-GLuc clone 8 was therefore removed from consideration for the MoN-Light BoNT assay.
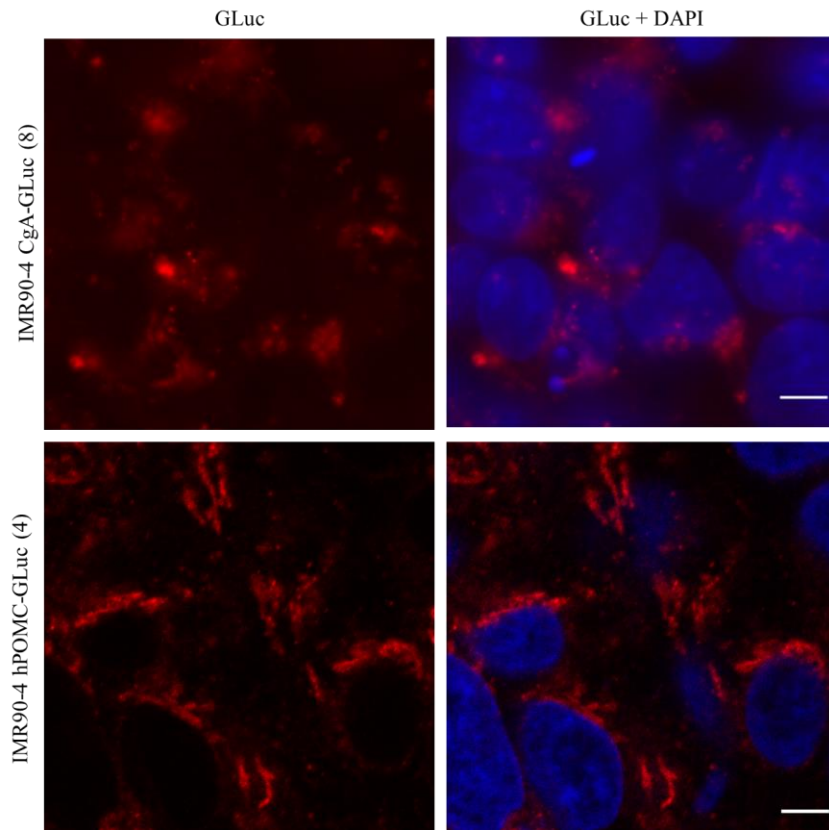
Figure 81: Immunofluorescence of GLuc and DAPI in IMR90-4 CgA-GLuc clone 8 and IMR90-4 hPOMC-GLuc 4. Cells were fixed, permeabilized and then incubated overnight with rabbit-α-GLuc and mouse-α-GM130 primary antibodies. Secondary antibody incubation with Alexa-fluor α-rabbit 568 (1:1000) and Alexa-fluor α-mouse 488 (1:1500). Cells were additionally incubated with DAPI for nuclei staining. scale bar = 5µm

IMR90-4 SgII-GLuc colocalization

The GLuc and GM130 proteins were identified by immunohistochemistry in the SgII-GLuc clone 11 (Figure 82). The appearance of GLuc after fluorescent labeling, contrary to the assessed expression levels (Figure 37) and the confirmed integration of the donor DNA into the AAVS1 safe harbor locus (Figure 60 and Figure 71), was extremely faint with no indication of sorting through the Golgi apparatus, therefore this clone was also removed from consideration for the MoN-Light BoNT assay.
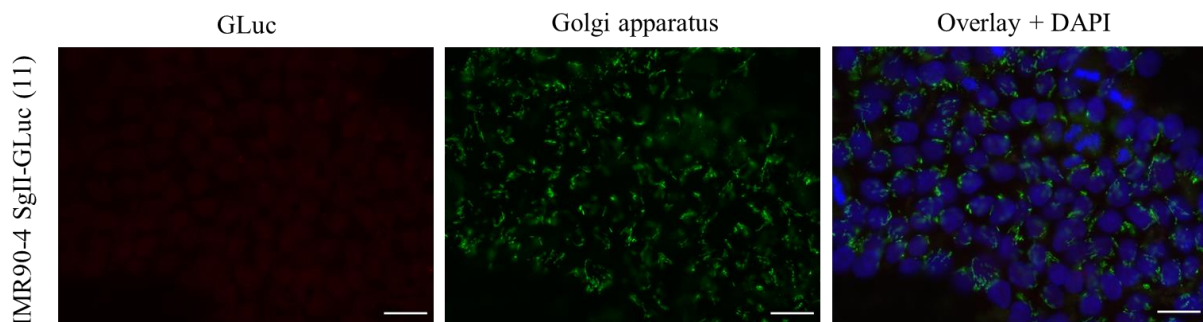
Figure 82: Immunofluorescence of GLuc and GM130 in IMR90-4 SgII clone. Cells were fixed, permeabilized and then incubated overnight with rabbit-α-GLuc and mouse-α-CgA primary antibodies. Secondary antibody incubation with Alexa-

128

Section 3.4

SIMA VAMP2-GLuc colocalization

The presence of GLuc in the SIMA VAMP2-GLuc clone 1 was analyzed for colocalization with Syp, which is a SV specific protein. The appearance of GLuc (green) after fluorescent labeling was diffuse, with little to no clustering. The labeled Syp protein (red) could be identified as clusters of protein to one side of the nucleus and staining in the axonal outgrowths from the soma. An overlay of these panels, including an image of the DAPI-stained nuclei contains arrows indicating the regions that were specifically analyzed for colocalization (Figure 83 A). There was little to no overlap in fluorescence levels between GLuc and Syp through the cell (Figure 83 B). The average overlap correlation score for these sections was 0.42 (Figure 83 C), even lower than the correlation score of the no tag GLuc clones analyzed. Because of the evident GLuc sorting failure, this clone group was excluded from further tests of luciferase localization or release in this project.
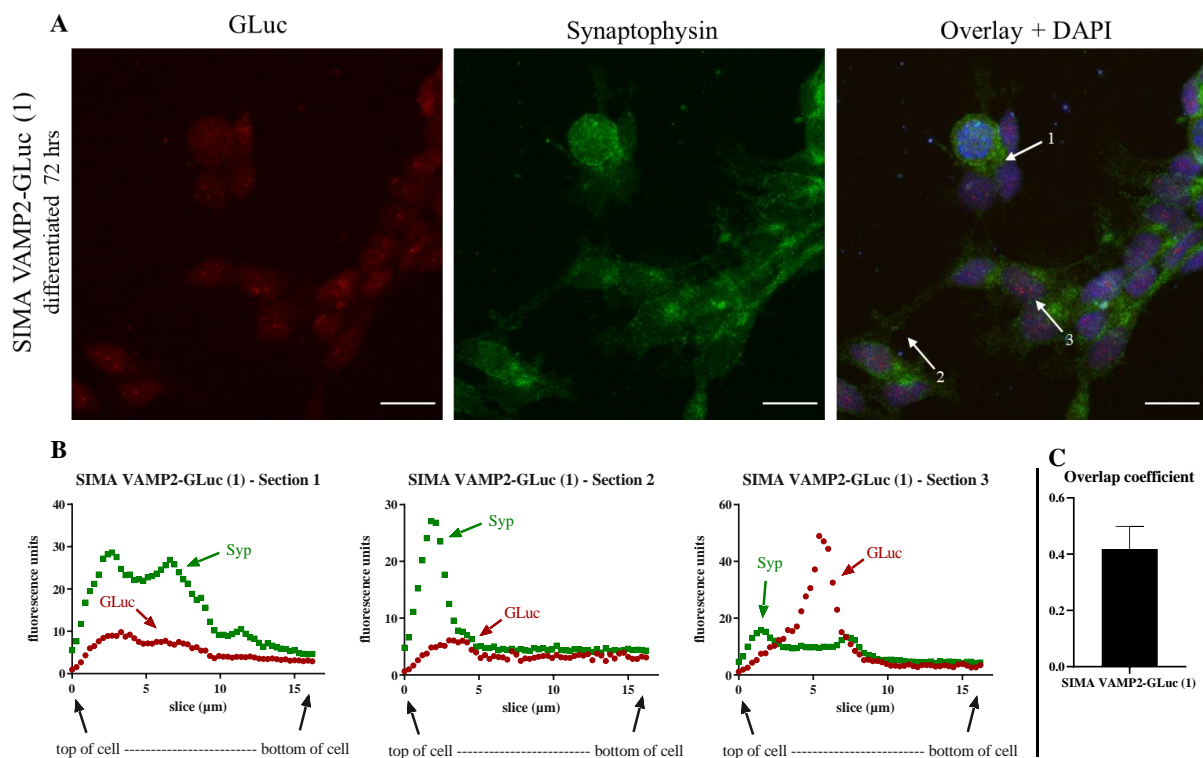
Figure 83: Colocalization of GLuc with Syp in SIMA VAMP2-GLuc clone. Upper panel A: immunofluorescence double labeling of SIMA VAMP2-GLuc-GLuc clone 1. Cells were fixed, permeabilized and then incubated overnight with rabbit-α-GLuc and mouse-α-Syp primary antibodies. Secondary antibody incubation with Alexa-fluor α-rabbit 568 (1:1000) and Alexa-fluor α-mouse 488 (1:1500). The overlay of all images includes DAPI labeled nuclei and sections of colocalization analysis indicated by white arrows in overlay image; scale bar = 20µm. Lower panel: (B) Graphical representation of colocalization analysis of GLuc with Syp by measurement of fluorescence units in selected cell sections in SIMA VAMP2-GLuc (1). (C) Overlap coeffient of GLuc and Syp colocalization.

IMR90-4 VAMP2-GLuc colocalization

The expression patterns of GLuc in IMR90-4 VAMP2-GLuc clone 11 (Figure 84 A) as visualized by immunohistochemistry looked quite different to that of GLuc in the SIMA VAMP2-GLuc clone. The IMR90-4 VAMP2-GLuc clone exhibited some clustering of GLuc (red) in specific locations in the cells, similar to, but more diffuse than, the hPOMC-GLuc clustering around the Golgi apparatus (green). The overlay includes the image of the DAPI-stained nuclei and the arrows indicating the regions that were more closely analyzed for GLuc and Golgi colocalization (Figure 84 A). The diagrams tracking fluorescence through the cell revealed generally lower fluorescence levels associated with GLuc, however, the fluorescence did rise in distinct peaks. The fluorescent levels associated with the Golgi apparatus also had a specific peak, as previously seen. While the intensities of each fluorescence level were not comparable, the peaks associated with GLuc localization were parallel to the peaks associated with the Golgi (Figure 84 B). The average overlap correlation score was 0.66 (Figure 84 C). The association between GLuc and the Golgi apparatus was only of a moderate degree, perhaps with only a fraction of the VAMP2-GLuc fusion protein passing through the secretory pathway.
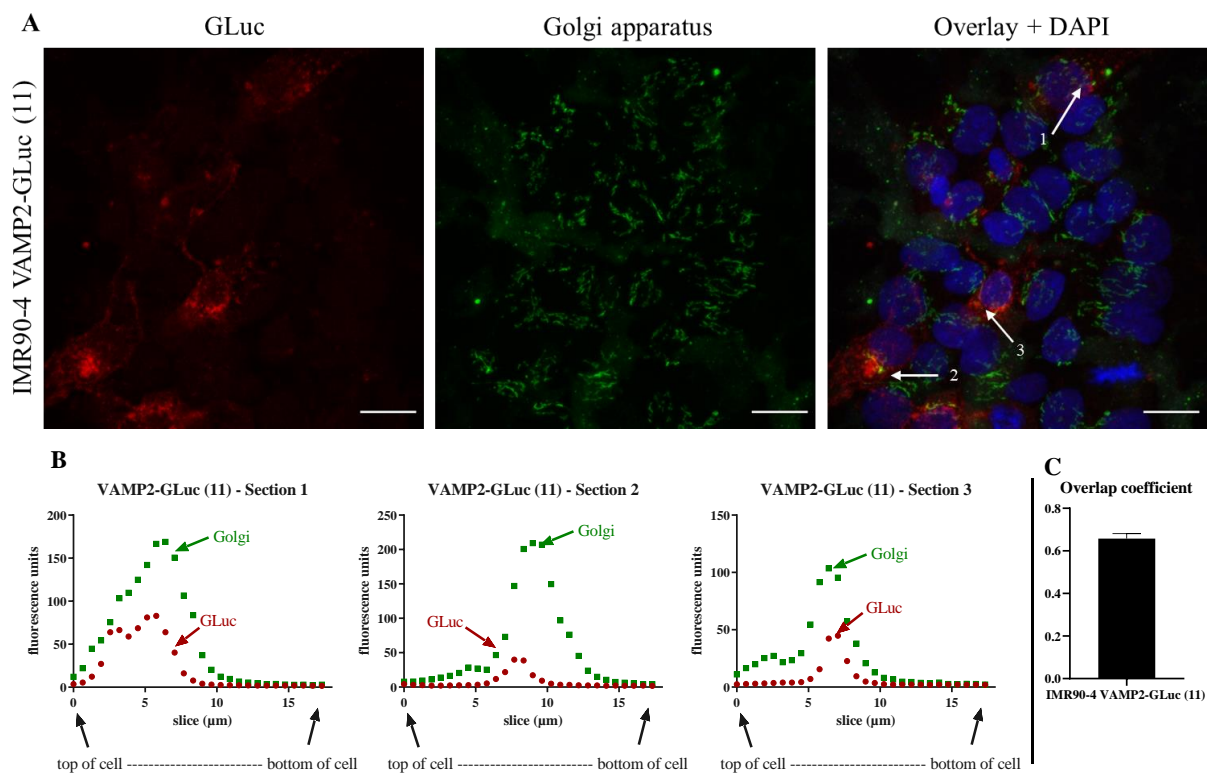
Figure 84: Colocalization of GLuc with Golgi in IMR90-4 VAMP2-GLuc clone. Upper panel A: immunofluorescence double labeling of IMR90-4 VAMP2-GLuc clone 11. Cells were fixed, permeabilized and then incubated overnight with rabbit-α-GLuc and mouse-α-CgA primary antibodies. Secondary antibody incubation with Alexa-fluor α-rabbit 568 (1:1000) and Alexa-fluor α-mouse 488 (1:1500). The overlay of all images includes DAPI labeled nuclei and sections of colocalization analysis indicated by white arrows in overlay image; scale bar = 20μm. Lower panel: (B) Graphical representation of colocalization analysis of

3.4.3.2   Summary of localization of signal peptide tagged GLuc

As summarized below, the colocalization of GLuc with the large dense core vesicle marker in the SIMA hPOMC-GLuc clone 4 was very high, as was that of GLuc with the Golgi marker in IMR90-4 hPOMC-GLuc clone 6. The colocalization of GLuc with the Golgi marker in IMR90-4 no tag GLuc clone 4 was quite low, but this was also expected, since the construct was designed to non-specifically express GLuc. The colocalization of GLuc with the Golgi marker in IMR90-4 VAMP2-GLuc clone 11 was only moderate, with a portion of GLuc passing through the Golgi, but another portion likely being missorted. On the other hand, the SIMA VAMP2-GLuc clone 1 was completely missorted and only resulted in a very low colocalization factor. Neither IMR90-4 CgA-GLuc clone 8 nor SgII-GLuc clone 11 can be used further, since GLuc appeared disrupted or only very faint in these clones (Figure 85).

	theoretical GLuc localization	insertion AAVS1 locus	off-target donor DNA insertions?	expected GLuc colocalization	
SIMA hPOMC-GLuc (clone 4)				CgA vs GLuc HIGH	✔
IMR90-4 hPOMC-GLuc (clone 6)				GM130 vs GLuc HIGH	✔
IMR90-4 no tag GLuc (clone 4)				GM130 vs GLuc LOW	✔
IMR90-4 CgA-GLuc (clone 8)				GM130 vs GLuc GLuc appears disrupted	✗
IMR90-4 SgII-GLuc (clone 11)				GM130 vs GLuc GLuc extremely faint	✗
SIMA VAMP2-GLuc (clone 1)				Syp vs GLuc LOW	✗
IMR90-4 VAMP2-GLuc (clone 11)				GM130 vs GLuc MEDIUM	✔

Figure 85: Continuation of summary of donor DNA integration and colocalization results of selected CRISPR-modified clones representing each construct.

3.4.4 Summary of clone characterization for MoN-Light BoNT assay

Insertion of donor DNA into AAVS1 safe harbor locus

The insert confirmation PCR provided evidence of homozygote insertions as well as the lack of insertions of the donor DNA at the AAVS1 safe harbor locus. There were no homozygote clones isolated from the SIMA transfections. The transfections resulting in IMR90-4 hPOMC-GLuc clones 4 and 6, IMR90-4 no tag GLuc clones 1 and 4, and IMR90-4 VAMP2-GLuc clone 11 contained homozygous insertions of the donor DNA at the AAVS1 safe harbor locus (Figure 86 A).

Identification of donor DNA integration events at AAVS1 safe harbor locus and beyond

The dc-qcnPCR analysis of donor DNA integration events provided evidence for integration events including a single integration of the donor DNA in one allele (IMR90-4 VAMP2-GLuc clone 16, Figure 73) to possibly hundreds of integration events (SIMA hPOMC-GLuc clone 29, Figure 67). IMR90-4 hPOMC-GLuc clones 4 and 6 and IMR90-4 VAMP2-GLuc clone 11 were demonstrated by qPCR analysis to carry two copies of the donor DNA. In combination with the insertion confirmation analyses, it was concluded that they were all homozygote clones with no off-target donor DNA integrations. The IMR90-4 no tag GLuc clones 1 and 4 were both homozygote clones, but were revealed to have at least couple off-target donor DNA integrations. The SIMA clones, including hPOMC-GLuc 1, 2, 4, and 5, and VAMP2-GLuc 1, were all heterozygote clones with at least one off-target integration. (Figure 86 B).

Colocalization of GLuc with markers of vesicular transport

The extent of colocalization, measured by the overlap coefficient, of GLuc and Golgi in both IMR90-4 hPOMC-GLuc clones 4 and 6 was significantly higher than that in the no tag GLuc clones 1 and 4. The degree of colocalization between GLuc and Golgi in the IMR90-4 VAMP2-GLuc clone 11 was only moderate. However, the IMR90-4 VAMP2-GLuc clone 11 still demonstrates a greater scale of colocalization between GLuc and GM130 in comparison to IMR90-4 no tag GLuc clone 4. The degree of colocalization between GLuc and Golgi in IMR90-4 VAMP2-GLuc clone 11 was also significantly higher than the degree of colocalization between GLuc and Syp in the SIMA VAMP2-GLuc clone 1. Furthermore, there was no significant difference between the degree of colocalization of the IMR90-4 no tag GLuc clones and the SIMA VAMP2-GLuc clone 1.
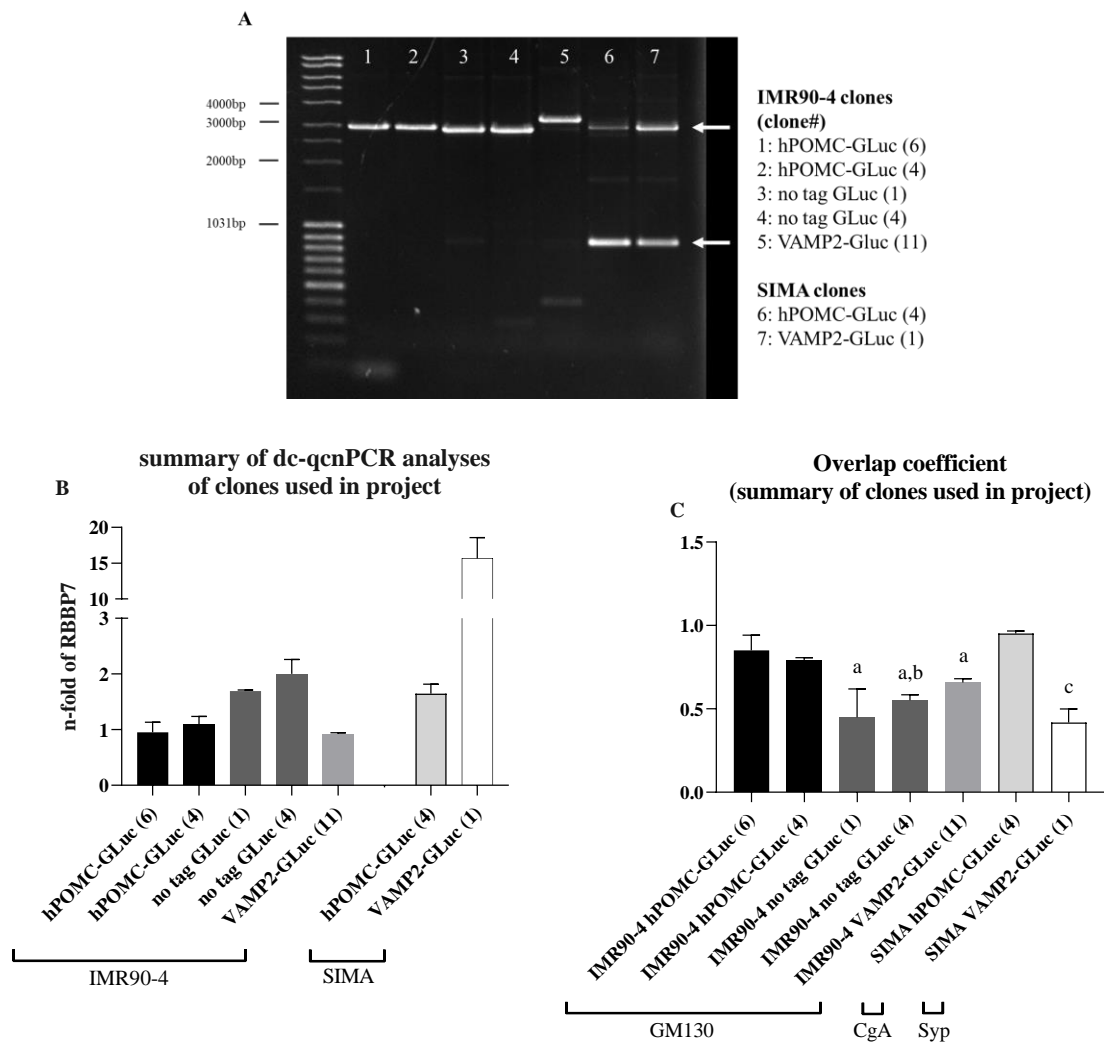
Figure 86: Characterization summary of selected CRISPR-modified clones. A: Donor DNA insert confirmation at AAVS1 safe harbor locus, summary of main CRISPR-modified clones used in this project. Expected product sizes: All WT alleles expected at 850bp, hPOMC-GLuc insert at 3000bp, no tag GLuc insert at 2900bp, VAMP2-GLuc insert at 3292bp. B: Summary of copy number analysis of GLuc, as compared to ChrX gene RBBP7 and normalized to autosomal gene CHOP. C: summary of degree of colocalization for all clones from sections 1-3, analysis with coloc2. Statistical differences measured by two-tailed t-test: a = clone vs IMR90-4 hPOMC-GLuc clones, $p < 0.05$; b = clone vs IMR90-4 VAMP2-GLuc, $p < 0.05$; c = clone vs IMR90-4 VAMP2-GLuc, $p < 0.001$.

The background for the selection of each characterized clone is summarized below. The SIMA hPOMC-GLuc clones 1, 2, 4, and 5 were all heterozygous clones with off-target donor DNA integrations, clone 4 was proven to sort a very high degree of GLuc into LDCVs. This clone was not intended for use in the MoN-Light BoNT assay, but rather as a screen during functional testing in the luciferase release assay before the IMR90-4 derived clones were differentiated into motor neurons. Due to the proven correct GLuc sorting, these clones were used in initial luciferase release tests, described below. The IMR90-4 hPOMC-GLuc clones 4 and 6 were both perfect candidates for use in the MoN-Light BoNT assay. IMR90-4 hPOMC-GLuc clone 6 was used for differentiation into motor neurons and functional testing described

134

below. The IMR90-4 no tag GLuc clones 1 and 4 contained both on and off-target donor DNA integrations, but due to the confirmed non-specific expression of GLuc IMR90-4 no tag GLuc clone 4 could be used as a control in the luciferase release assay. The IMR90-4 CgA-GLuc, IMR90-4 SgII-GLuc and SIMA VAMP2-GLuc clones did not pass the full characterization tests and were not used for any functional testing. Despite the only moderate degree of sorting of GLuc through the Golgi in the IMR90-4 VAMP2-GLuc clone 11, this construct was pursued for its potential usefulness in the MoN-Light BoNT assay.

	theoretical GLuc localization	insertion AAVS1 locus	off-target donor DNA insertions?	Expected GLuc colocalization	Clone used for functional testing?
SIMA hPOMC-GLuc (clones 1, 2, 4, 5)					✔
IMR90-4 hPOMC-GLuc (clone 6)					✔
IMR90-4 no tag GLuc (clone 4)					✔
IMR90-4 CgA-GLuc (clone 8)					✘
IMR90-4 SgII-GLuc (clone 11)					✘
SIMA VAMP2-GLuc (clone 1)					✘
IMR90-4 VAMP2-GLuc (clone 11)					✔

Figure 87: Final summary of the characterization of selected CRISPR-modified clones for each construct prepared and analyzed in this project.

3.5 Luciferase release upon cellular depolarization

3.5.1 SIMA hPOMC-GLuc luciferase release and detection

Previously, it was shown that a depolarization dependent release of luciferase can be inhibited by exposure to botulinum neurotoxin from the SIMA Random-Insertion_hPOMC1-26GLuc prototype that contains multiple copies of hPOMC-GLuc randomly inserted into the genome[2]. Four unique CRISPR-modified SIMA hPOMC-GLuc clones (1, 2, 4, and 5) were used to test whether the insertion of hPOMC-GLuc in the AAVS1 safe harbor locus by CRISPR/Cas is able to produce a measurable release of luciferase upon cellular depolarization. Ideally, these clones should have no off-target donor DNA integrations, however no CRISPR-modified SIMA clones were identified to have only on-target integrations. Therefore, luciferase release was tested despite the fact the clones did not meet the MoN-Light BoNT quality criteria.

The four clones were exposed to Na^+-HBS control buffer or K^+-HBS depolarization buffer for 3 minutes at 37 °C. The buffer was collected and measured for luciferase activity in the supernatant. In all clones, the release of luciferase into the supernatant was increased by depolarization with K^+-HBS buffer. Although clones 4 and 5 contained fewer off-target copies of the hPOMC-GLuc coding sequence and clones 1 and 2 contained greater number of off-target hPOMC-GLuc insertions, clones 2 and 4 released higher activities of GLuc than clones 1 and 5, both under non-depolarizing and depolarizing conditions (Figure 88 A). The ratio of release with and without depolarization was, however, identical in all clones (Figure 88 B). The increase of measured luciferase in the supernatant of depolarized SIMA Random-Insertion_hPOMC1-26GLuc prototype cells also reaches just over 3-fold that of the cells in the control buffer[2].
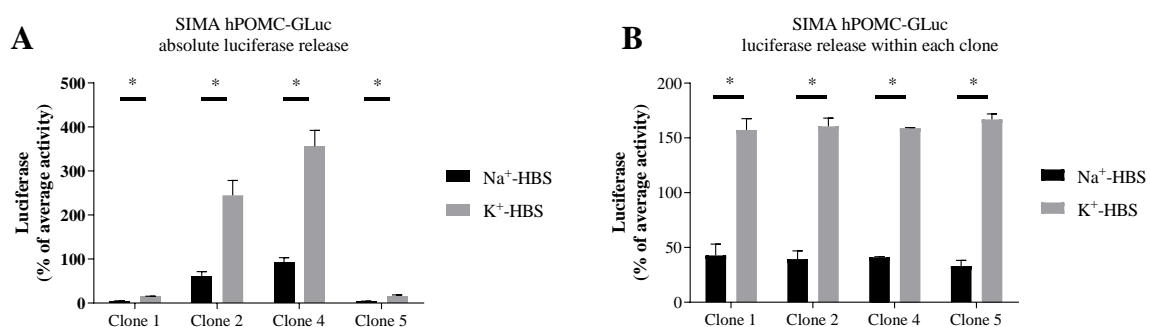
Figure 88: Luciferase activity measured from four unique SIMA hPOMC-GLuc clones. Cells were differentiated for 72 hours prior to undergoing release assay. Cells were pre-incubated with fresh medium for 10 min at 37°C. Release of luciferase activity was stimulated with control (Na^+-HBS) or stimulation buffer (K^+-HBS) for 3 min at 37°C and luciferase activity in the supernatant was measured. A: The average of the luciferase activity in all conditions and all clones was calculated and the percent difference from the average for each condition is presented as means ± SD of 9 measurements in three independent

136

To test whether the depolarization-dependent release occurred specifically through the calcium-mediated fusion of neurosecretory vesicles to the presynaptic membrane, the calcium chelator EGTA was added to the luciferase release experiment. Two SIMA hPOMC-GLuc clones (1 and 4) were examined in parallel. The depolarization of the cells under normal conditions resulted in an approximate 3-fold increase in luciferase activity in the supernatant. The addition of 10 mM EGTA to the buffers completely inhibited the depolarization-dependent release of luciferase, while it did not affect the depolarization-independent background luciferase activity detected in presence of Na^+-HBS (Figure 89).
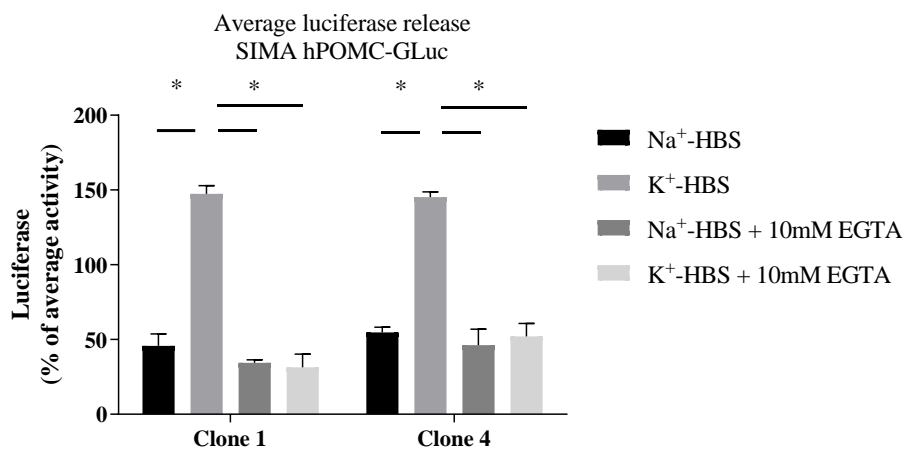
Figure 89: Luciferase activity measured from two separate SIMA hPOMC-GLuc clones. Cells were differentiated for 72 hours prior to undergoing release assay. Cells were pre-incubated with fresh medium for 10 min at 37°C. Release of luciferase activity was stimulated with control (Na^+-HBS) or stimulation buffer (K^+-HBS) in the absence or presence of 10 mM EGTA for 3 min at 37°C and luciferase activity in the supernatant was measured. The average of the luciferase activity in Na^+-HBS and K^+-HBS supernatants from each individual clone was calculated and the percent difference from the average for each condition is presented as means ± SD of 9 measurements in three independent experiments. * = p < 0.01

3.6   Establishment of motor neuron differentiation protocols

Three protocols for the differentiation of iPSCs into motor neurons were compared (Du *et al*[8], Maury *et al*[9], and Kroehne *et al*[111]). The investigation of the efficacy of these protocols was carried out by Maren Schenke at the University of Veterinary Medicine in Hannover and a very short summary of her findings is described below. A summary of the steps involved in carrying out each of the three differentiation protocols is illustrated in Figure 90.
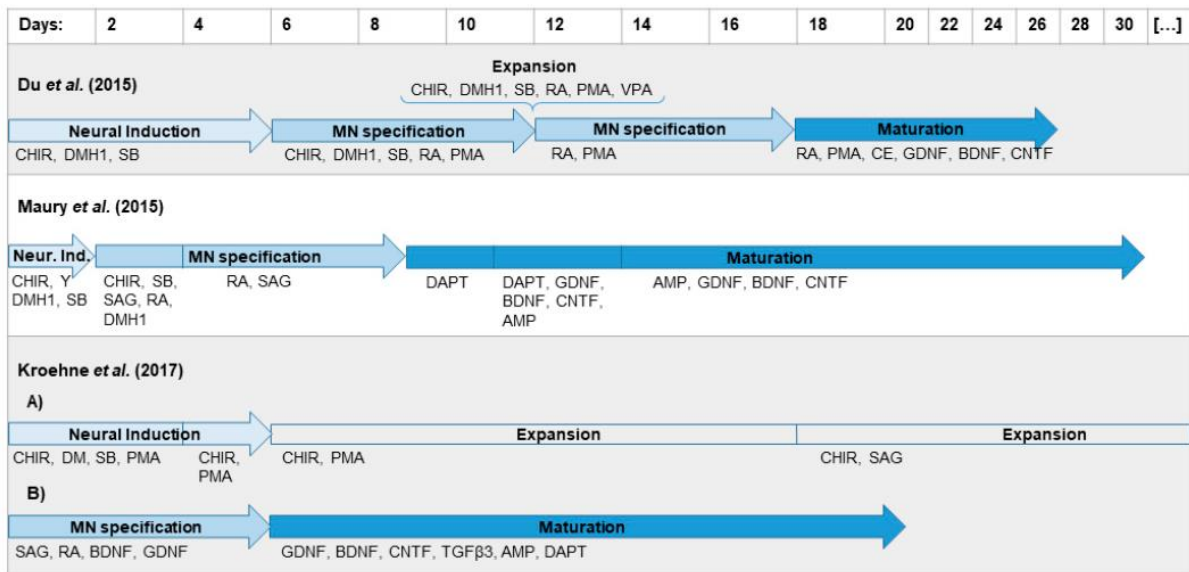
Figure 90: Summary of differentiation protocols from iPSC to MN states. Figure published in Schenke et al[99], licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

According to the work of Schenke *et al*[99], the yield of MNs using the protocol described by Du *et al* reached an average of 50 % of the cell population but this number varied greatly between experiments. Despite the high variation, the MN population derived from the Du *et al* protocol was significantly higher than the MN yield using protocols based on the work of Maury *et al* and Kroehne *et al*, both of which only reached approximately 15 % (Figure 91).
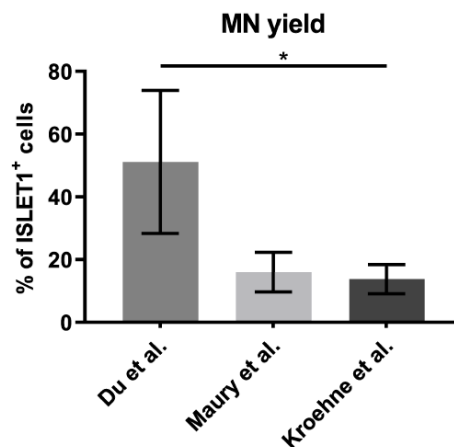
*Figure 91: MN yield as measured by percentage of Islet1 cells on day 28 of the protocol based on Du[8], day 32 of the protocol based on Maury[9], and day 21 of the protocol based on Kroehne[111]. * $p < 0.05$ Figure published in Schenke et al99, licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).*

Importantly, further analyses revealed that all three differentiation protocols produced MNs expressing the most important BoNT receptors and targets, such as the SV2, Syt, and gangliosides important for uptake into the cell and the members of the SNARE complex (Figure 92). Therefore, theoretically any of the three differentiation methods could be used to obtain cells that are sensitive to BoNT.

138

Section 3.6
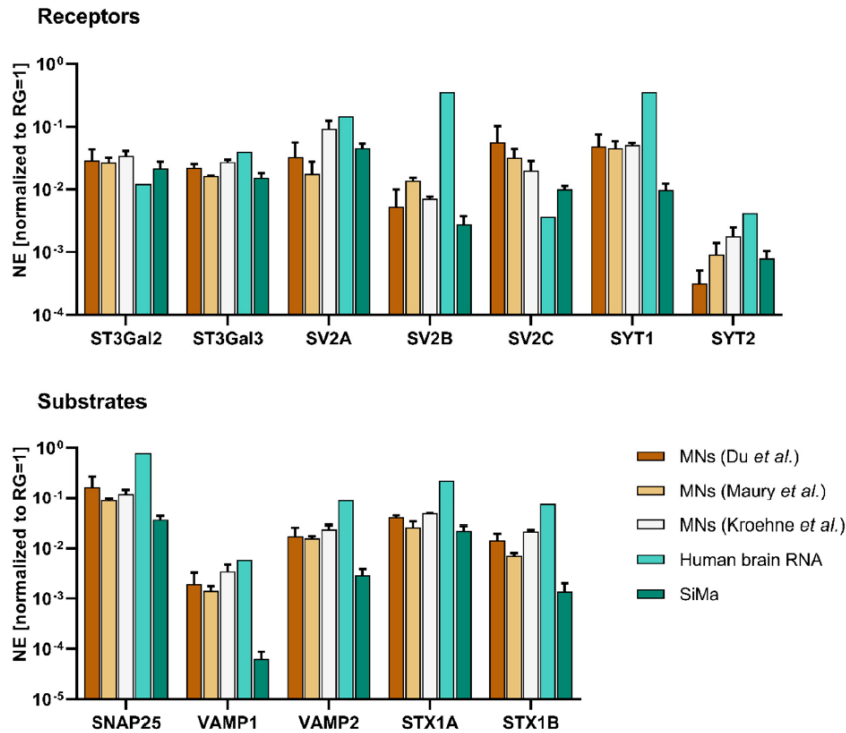
Figure 92: Gene expression levels of BoNT receptors and targets in MNs differentiated with the three test protocols. Figure published in Schenke et al99, licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

3.7   Validation of IMR90-4 pluripotent status

The genes used in this project to verify the pluripotent state of the IMR90-4 cells are *LIN28*, *NANOG*, *OCT4* and *SOX2*. These four factors were originally used to reprogram IMR90-4 cells into iPSCs and should therefore be expressed in these cells pre-differentiation. The expression of these four markers in non-transfected IMR90-4 cells, the two IMR90-4 hPOMC-GLuc clones 4 and 6, and the IMR90-4 VAMP2-GLuc clone 11 were compared with their expression in hPOMC-GLuc MNs, differentiated according to Maury *et al* for 30 days (Figure 93). All four markers were expressed in the four iPSC cell lines, but not in the MNs. This verifies the iPSC status of these clones.
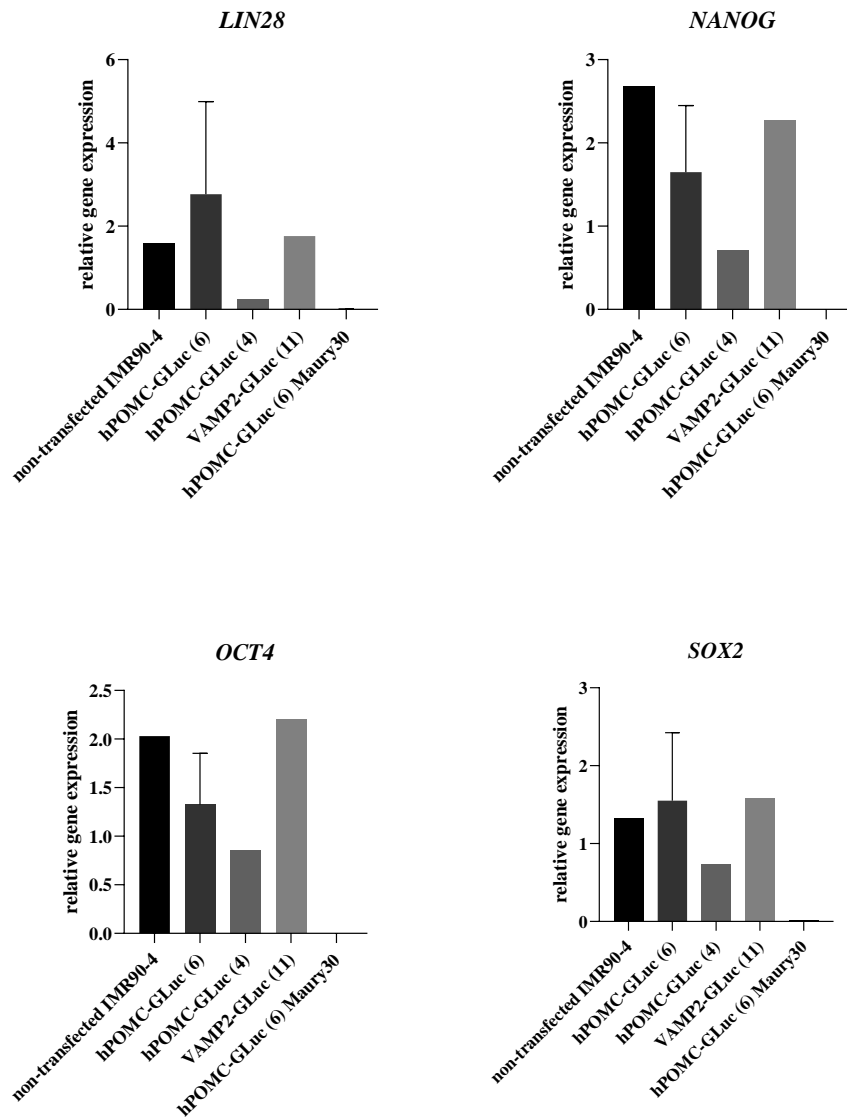
Figure 93: iPSC status of pre-differentiated non-transfected IMR90-4 and selected genetically modified clones by expression analysis of iPSC reprograming markers. Relative gene expression in iPSCs was compared to expression in hPOMC-GLuc clone 6 after 30 days of differentiation and calculated by the ΔΔCt method, with the genes RPS23 and PPIA used for normalization. n = 1-3 biological replicates.

3.8   Differentiation of induced pluripotent stem cells into motor neurons

The differentiation protocol based on Maury *et al* is the easiest protocol to execute and would therefore be the most suitable protocol for the final purpose, the establishment of an easily usable BoNT potency assay. For this reason, the most complete data in the following section is from MNs differentiated following a protocol adapted from Maury *et al*. Because the protocol proposed by Du *et al* yielded a higher number of motor neurons this protocol was also implemented. Motor neuron differentiations were performed with the IMR90-4 hPOMC-GLuc

140

clone 6, the IMR90-4 VAMP2-GLuc clone 11, and, for comparison, the IMR90-4 no tag GLuc clone 4. The differentiation based upon Maury *et al* could be accomplished in all three clones, producing Islet1 positive motor neurons. However, it was not possible to obtain hPOMC-GLuc motor neurons according to Du *et al*. All cells died between days six and twelve of the differentiation period. Using the same protocol, the VAMP2-GLuc cells could be differentiated into motor neurons.

3.8.1 Expression analysis of GLuc and MN genes

After 30 days of exposure to differentiation medium (see Material and Methods 2.3.3), the presence of motor neurons and the expression of GLuc in the resulting population was verified. qPCR analysis was carried out to compare the expression of various genes in cells differentiated according to Maury *et al* and Du *et al* protocols compared to their corresponding non-differentiated iPSCs (Figure 94). Neither hPOMC-GLuc nor VAMP2-GLuc were detectable in non-transfected IMR90-4 cells (Appendix 6.7, Figure 149). During the differentiation protocol according to Maury *et al*, the expression level of hPOMC-GLuc decreased about 2-fold in comparison to non-differentiated IMR90-4 hPOMC-GLuc cells of the same clone. Even more dramatically, the expression of VAMP2-GLuc in cells differentiated according to Maury *et al* decreased by about 20-fold. In contrast, in IMR90-4 VAMP2-GLuc cells that were differentiated according to Du *et al* the expression of GLuc only decreased by approximately 4-fold.

The expression of key genes associated with the successful differentiation into motor neurons, *CHAT*, *ISLET*, and *HB9*, was measured. All the iPSCs included in the expression analysis apparently expressed low levels of each motor neuron marker, however in comparison to the much higher expression of the genes in most of the differentiated cells the expression in non-differentiated cells is negligible. The IMR90-4 hPOMC-GLuc clone differentiated according to Maury *et al* expressed significantly higher amounts of all genes associated with successful motor neuron differentiation in comparison to its non-differentiated counterpart. The IMR90-4 VAMP2-GLuc clone differentiated according to both Maury *et al* and Du *et al* also expressed significantly higher amounts of *CHAT*, *ISLET*, and *HB9* in comparison to non-differentiated IMR90-4 VAMP2-GLuc. The IMR90-4 no tag GLuc clone differentiated according to Maury *et al* expressed very low quantities of *HB9* and *CHAT*, but high amounts of *ISLET1* (Figure 95).

Change in expression of key genes in clones before and after motor neuron differentiation
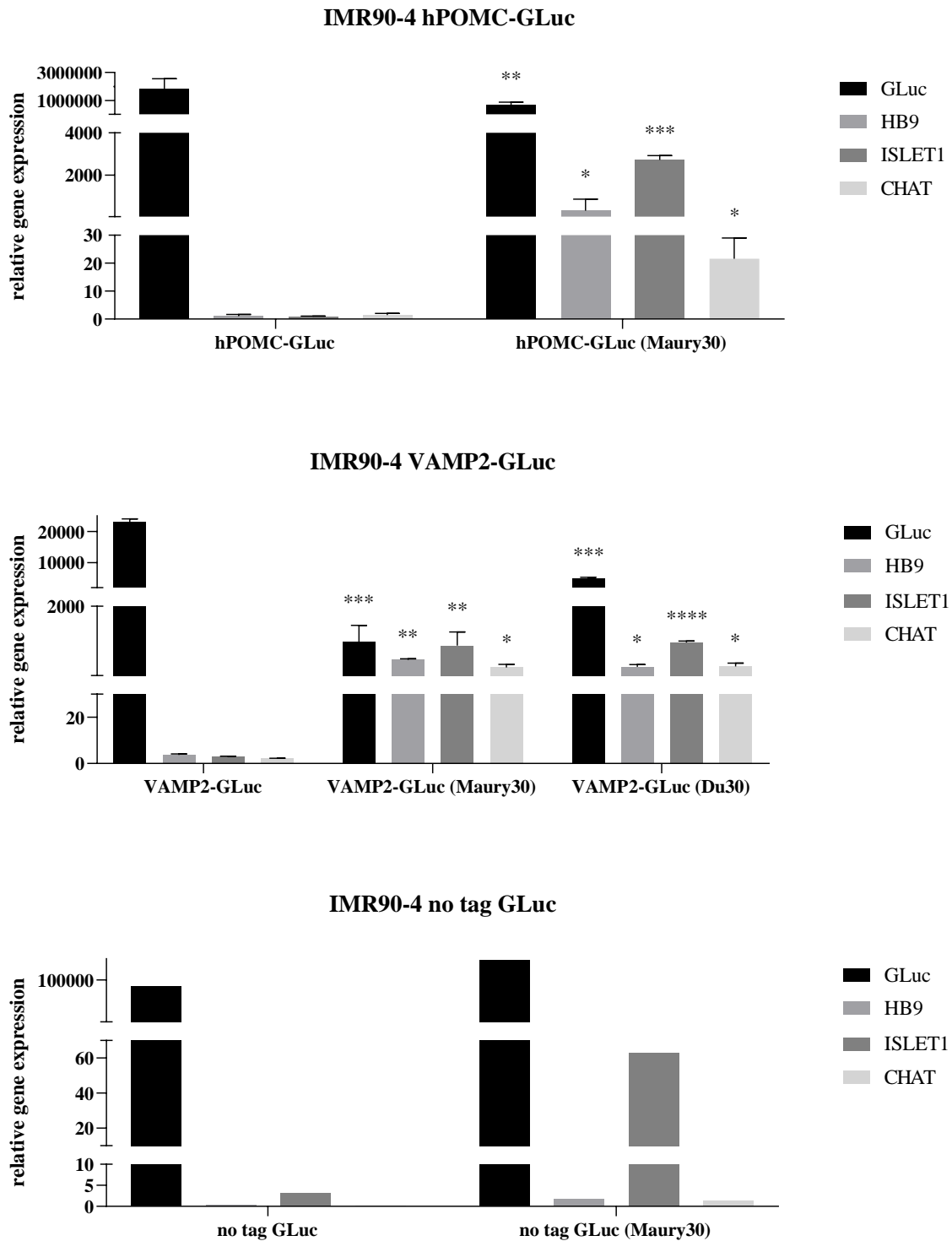
Figure 94: GLuc expression in CRISPR-modified iPSCs and in day 30 Maury and Du differentiated clones. Differentiation status of differentiated clones assessed by relative gene expression of selected genes associated with MN status and calculated by the $\Delta\Delta Ct$ method, with the genes RPS23 and PPIA used for normalization. Statistical significance measured by t-test, change in expression levels between differentiated clone versus clone in pluripotent state. Statistical significance evaluated with t-test * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, *** = $p < 0.0001$. IMR90-4 hPOMC-GLuc and IMR90-4 VAMP2-

3.9 Immunofluorescence post-differentiation

Immunofluorescence was carried out in order to verify expression of GLuc in differentiated cells with a second independent experiment. This method was also used to provide evidence of the co-expression of motor neuron marker Islet1 with GLuc in differentiated cells. Furthermore, it was used to verify the colocalization of GLuc with large dense core vesicles or synaptic vesicles in IMR90-4 hPOMC-GLuc clone 6 and IMKR90-4 VAMP2-GLuc clone 11, respectively.

3.9.1 Verification of GLuc expression in differentiated cells

GLuc could be detected by immunofluorescence in IMR90-4 hPOMC-GLuc clone 6 cells differentiated for 30 days according to Maury *et al*. Furthermore, GLuc could be detected in IMR90-4 VAMP2-GLuc clone 11 cells on day 30 of both Maury *et al* and Du *et al* differentiation protocols. In all three panels GLuc was found in areas surrounding and in processes extending from the soma (Figure 95). Several nuclei are stained for DAPI, but not for GLuc. In all differentiation protocols, it was observed that multiple apoptotic cells, identified by condensed and brightly stained nuclei[112], often surround the live neuronal cells. Depending on the time of apoptosis, GLuc may no longer be found in these cells. Although some cells may also have lost GLuc expression during the differentiation process.
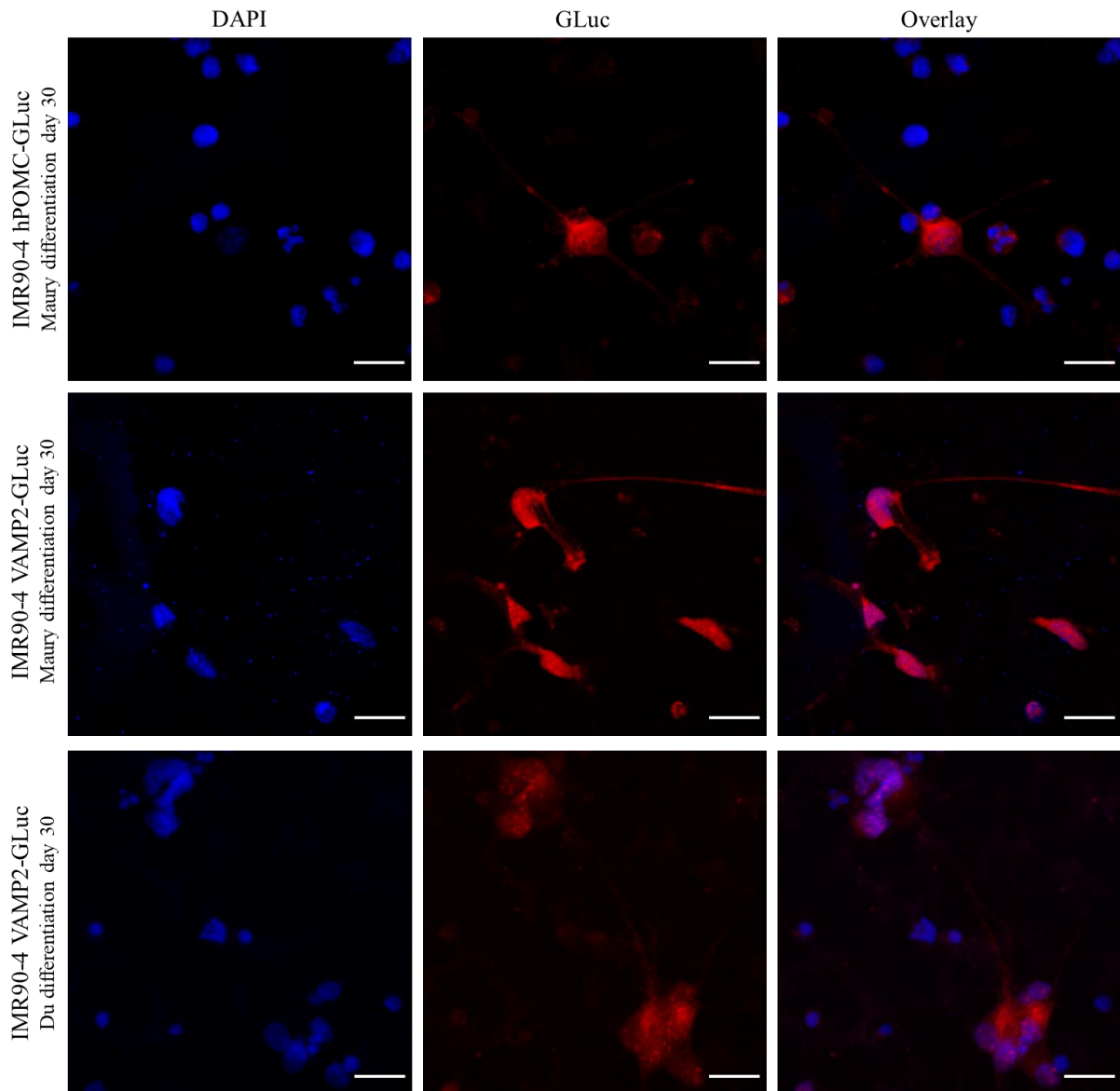
Figure 95: GLuc expression in MNs differentiated for 30 days. Top panel: IMR90-4 hPOMC GLuc clone differentiated according to Maury et al and fixed on day 30; Middle panel: IMR90-4 VAMP2-GLuc clone differentiated according to Maury et al and fixed on day 30; Bottom panel: IMR90-4 VAMP2-GLuc clone differentiated according to Du et al and fixed on day 30. All panels show cells labeled with DAPI and GLuc, plus the overlay. Cells were fixed, permeabilized and then incubated overnight with rabbit-α-GLuc primary antibody. Secondary antibody incubation with Alexa-fluor α-rabbit 568 (1:1000). scale bar = 20 µm

3.9.2 Confirmation of GLuc co-expression with motor neuron marker Islet1

qPCR analyses indicated that at least a proportion of the resulting neuronal population consisted of motor neurons expressing *ISLET1* and GLuc (Figure 94). Immunofluorescence and confocal microscopy were carried out to verify the co-expression of GLuc and Islet1 in individual cells.

IMR90-4 hPOMC-GLuc clones differentiated according to Maury et al

IMR90-4 hPOMC-GLuc cells were differentiated according to the Maury *et al* protocol for 30 days. Several cells were included in the sample image, as seen in the DAPI staining. Only one of these cells also was labeled positively as Islet1 (green), indicating that only a proportion of the cells in this population are actually motor neurons. However, as seen in the overlay of images identifying Islet1, GLuc, and DAPI, GLuc is expressed in the same cell as Islet1 (Figure 96).
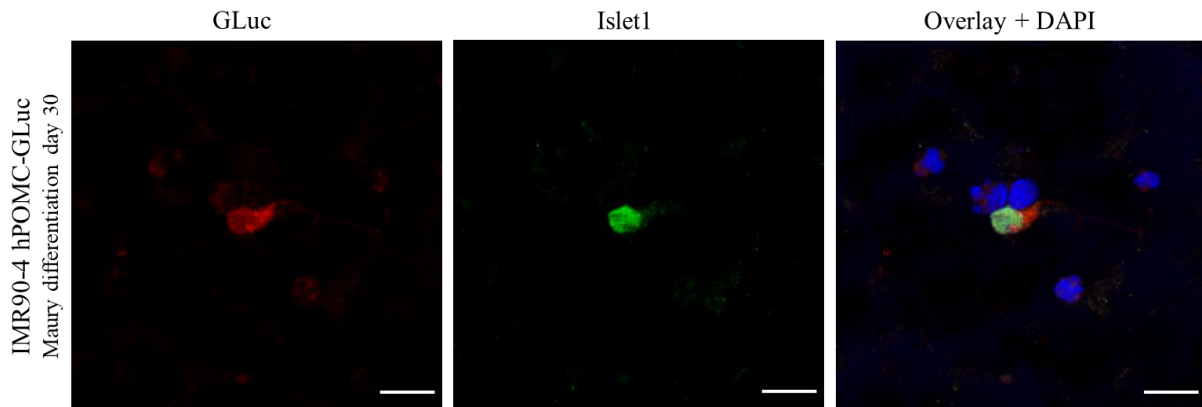
Figure 96: Co-expression of Islet1 and GLuc in IMR90-4 hPOMC-GLuc cell (Maury D30), fixed on day 30. Cells were fixed, permeabilized and then incubated overnight with rabbit-α-GLuc and mouse-α-Islet1 [1H9] primary antibodies. Secondary antibody incubation with Alexa-fluor α-rabbit 568 (1:1000) and Alexa-fluor α-mouse 488 (1:1500). Cells were additionally incubated with DAPI for nuclei staining. scale bar = 20 μm

IMR90-4 VAMP2-GLuc clones differentiated according to Maury et al

A similar pattern was seen in IMR90-4 VAMP2-GLuc cells differentiated according to Maury *et al* for 30 days. In the image of DAPI stained cells, several apoptotic cells were seen next to a single live cell. This same live cell was identified by immunohistochemistry to express both Islet1 (green) and GLuc (red). The overlay of all images confirmed that all protein labelling occurred in the same cell ensuring the co-expression of GLuc and Islet1 (Figure 97).
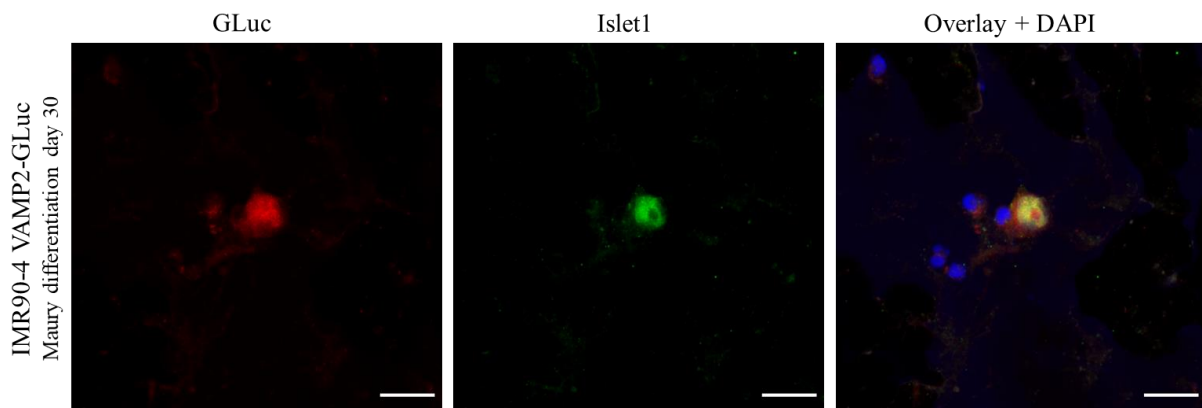
Figure 97: Co-expression of Islet1 and GLuc in an IMR90-4 VAMP2-GLuc cell (Maury D30). Cells were fixed, permeabilized and then incubated overnight with rabbit-α-GLuc and mouse-α-Islet1 [1H9] primary antibodies. Secondary antibody incubation with Alexa-fluor α-rabbit 568 (1:1000) and Alexa-fluor α-mouse 488 (1:1500). Cells were additionally incubated with DAPI for nuclei staining. scale bar = 20 μm

145

Section 3.9

IMR90-4 VAMP2-GLuc cells were also differentiated according to Du *et al* for 30 days. Multiple cells were stained with DAPI (blue channel), several of which were also identified to express Islet1 (green channel). These same Islet1 labeled cells expressed GLuc (red channel). The colocalization was confirmed by the image overlay (Figure 98). These analyses confirmed that for both IMR90-4 hPOMC-GLuc and IMR90-4 VAMP2-GLuc clones, the GLuc protein was expressed in Islet1 positive motor neurons.
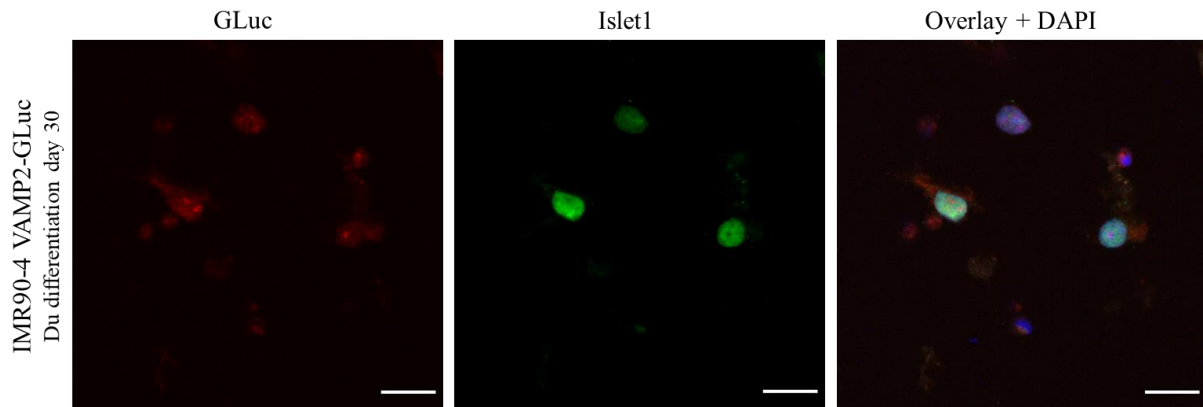
Figure 98: Co-expression of Islet1 and GLuc in IMR90-4 VAMP2-GLuc cells (Du D30) fixed on day 30. Cells were fixed, permeabilized and then incubated overnight with rabbit-α-GLuc and mouse-α-Islet1 [1H9] primary antibodies. Secondary antibody incubation with Alexa-fluor α-rabbit 568 (1:1000) and Alexa-fluor α-mouse 488 (1:1500). Cells were additionally incubated with DAPI for nuclei staining. scale bar = 20 μm

These analyses indicated that the successful depolarization of the motor neurons should concurrently release GLuc, in the case of IMR90-4 hPOMC-GLuc, or expose GLuc, in the case of IMR90-4 VAMP2-GLuc, to the supernatant surrounding the cells. Therefore there should be a measurable increase in luciferase activity in the supernatant if neurosecretory vesicle fusion with the presynaptic terminal membrane takes place. Although the number of motor neurons obtained in the different protocols varied and the expression level of the different GLuc constructs differed, it appeared that all motor neurons expressed the transgene. Thus, the hypothesis that the decrease in total GLuc expression after subjecting the cells to the differentiation protocols was the consequence of a complete loss of expression in differentiated cells was refuted.

3.9.3 Colocalization of GLuc with large dense core vesicles or synaptic vesicles

Finally, it is necessary to confirm that the colocalization of GLuc with the Golgi apparatus in the undifferentiated IMR90-4 clones remained consistent and therefore transforms into a colocalization of GLuc with large dense core vesicles or synaptic vesicles in hPOMC-GLuc and VAMP2-GLuc clones, respectively. Colocalization was analyzed by confocal

146

immunofluorescence microscopy (as described in section 2.14.3) at the end of the differentiation period.

IMR90-4 hPOMC-GLuc clone differentiated according to Maury et al

Colocalization of GLuc and the LDCV-associated protein SgII was carried out in the IMR90-4 hPOMC-GLuc clone 6 differentiated for 30 days according to Maury *et al*. Immunofluorescence identifying GLuc (red channel, Figure 99 A) was more widespread through the cell than SgII immunolabeling (green channel, Figure 99 A), which was restricted to an accumulation next to the nucleus and small aggregates extended along extruding processes. It was visibly evident in the overlay image (Figure 99 A) that in some places both proteins were expressed in the same cellular locations. Regions for colocalization analysis were selected from putative vesicles traveling in processes extending from the soma (arrows 1 and 2 in the overlay) and a region highly expressing SgII adjacent to the nucleus (arrow 3 in the overlay). While the SgII associated fluorescence intensity (green) was slightly lower than the GLuc associated fluorescence intensity (red), the Z-axis plots clearly delineated a colocalization of GLuc and SgII, with the protein traces running parallel to one another and peaking at the same point in the cell (Figure 99 B). The voxel analysis of colocalization resulted in an average overlap correlation score of 0.85 (Figure 99 C). The colocalization of GLuc and SgII was confirmed by the fluorescence analysis and the correlation score, indicating that GLuc is successfully sorted into LDCVs with SgII.
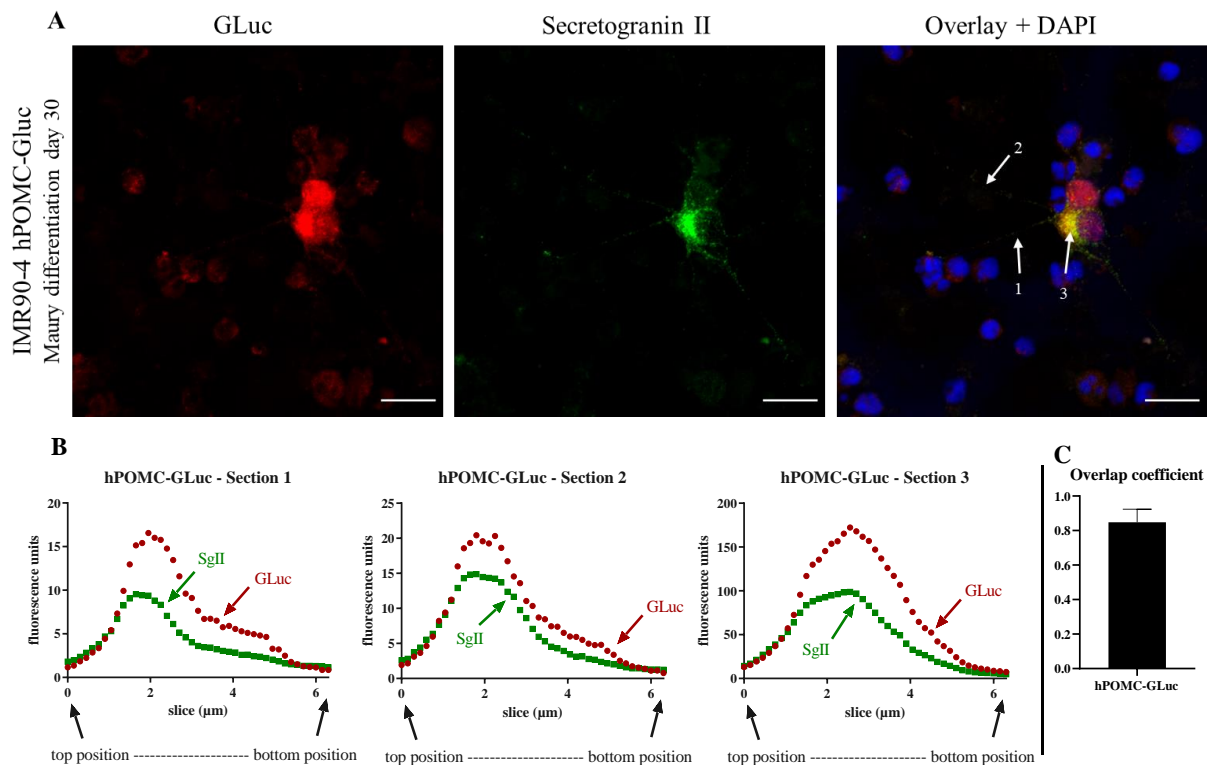
Figure 99: Colocalization analysis of GLuc and SgII in IMR90-4 hPOMC-GLuc MNs (Maury D30) protocol. Upper panel A: confocal images of immunofluorescence double labeling of IMR90-4 hPOMC-GLuc clone 6. Cells were fixed, permeabilized and then incubated overnight with rabbit-α-GLuc and mouse-α-SgII primary antibodies. Secondary antibody incubation with Alexa-fluor α-rabbit 568 (1:1000) and Alexa-fluor α-mouse 488 (1:1500). The overlay of all images includes DAPI labeled nuclei and sections of colocalization analysis indicated by white arrows in overlay image; scale bar = 20μm. Bottom panel: (B) z-axis fluorescence analysis of sections indicated in the overlay image, tracing the localization of GLuc and SgII. (C) Overlap coefficient for colocalization of GLuc with SgII.

IMR90-4 VAMP2-GLuc clone differentiated according to Maury et al

Colocalization of GLuc and the SV-associated protein Syp was carried out in the IMR90-4 VAMP2-GLuc clone 11 differentiated for 30 days according to Maury *et al*. Both GLuc and Syp were identified by immunohistochemistry in the processes extending from the soma, as well as in the area surrounding the cell nucleus (Figure 100 A). Regions for colocalization analysis were selected from Syp-positive vesicles traveling in processes extending from the soma (arrows 1 and 2 in the overlay) and an area adjacent to the nucleus (arrow 3 in the overlay). The fluorescence levels associated with GLuc (red) and Syp (green) in sections 1 and 2 rose in parallel. On the other hand, the peaks of fluorescence intensity for each protein in section 3 occurred in two different positions of the cell (Figure 100 B). The average overlap correlation score for these sections was 0.65 (Figure 100 C). Colocalization of the VAMP2-GLuc construct with Syp into the SVs was confirmed, but the mismatched peaks in section 3 also indicated a degree of mis-sorting.

148
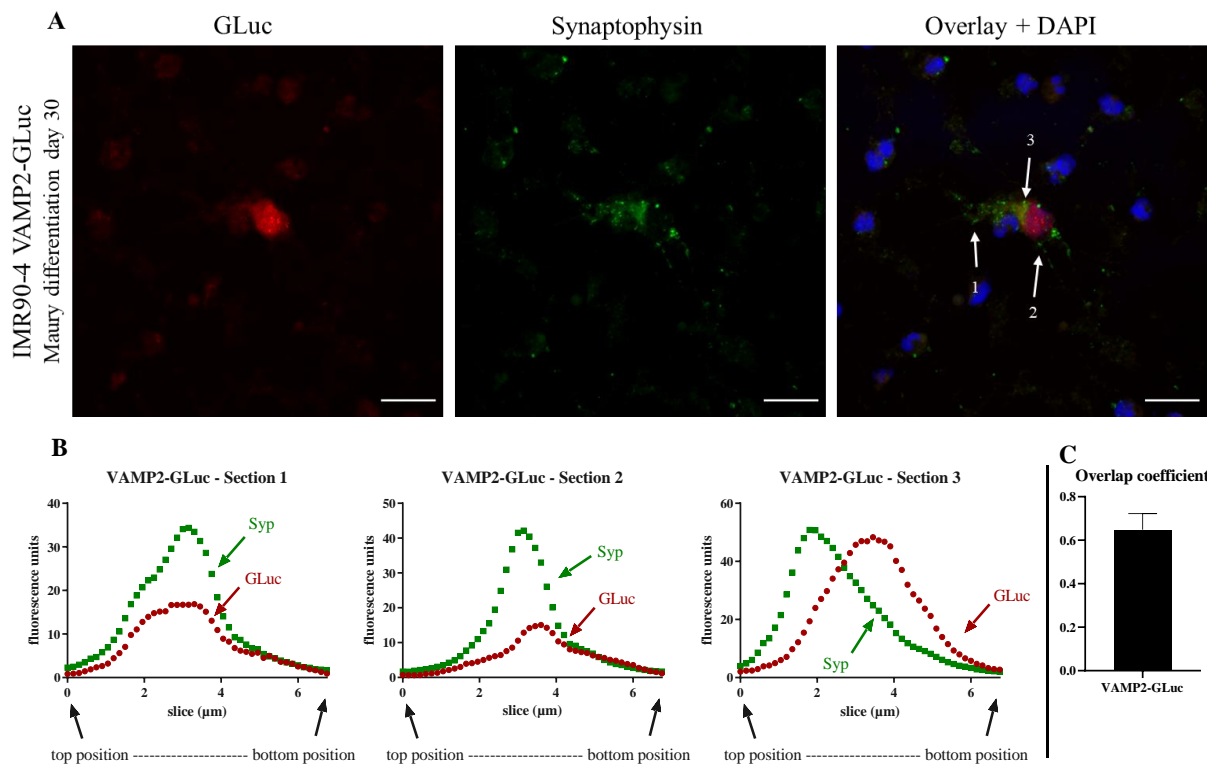
Figure 100: Colocalization analysis of GLuc and Syp in IMR90-4 VAMP2-GLuc MNs (Maury D30) protocol. Upper panel A: confocal images of immunofluorescence double labeling of IMR90-4 VAMP2-GLuc clone 11. Cells were fixed, permeabilized and then incubated overnight with rabbit-α-GLuc and mouse-α-Syp primary antibodies. Secondary antibody incubation with Alexa-fluor α-rabbit 568 (1:1000) and Alexa-fluor α-mouse 488 (1:1500). The overlay of all images includes DAPI labeled nuclei and sections of colocalization analysis indicated by white arrows in overlay image; scale bar = 20µm. Bottom panel: (B) z-axis fluorescence analysis of sections indicated in overly image, tracing the localization of GLuc and Syp. (C) Overlap coefficient for colocalization of GLuc with Syp.

IMR90-4 VAMP2-GLuc differentiated according to Du et al

Colocalization of GLuc and the SV-associated protein Syp was carried out in IMR90-4 VAMP2-GLuc clone 11 differentiated for 30 days according to Du *et al*. Both GLuc (red channel, Figure 101 A) and Syp (green channel, Figure 101 A) were identified with immunofluorescence in areas surrounding the nucleus and processes extending from the soma. Regions for colocalization analysis were selected from Syp-positive vesicles traveling in processes extending from the soma (arrows 1 and 2 in the overlay) and an area adjacent to the nucleus (arrow 3 in the overlay). The Z-axis plots of fluorescence intensities associated with GLuc (red) and Syp (green) overlapped in all three sections with peaks in protein localization occurring at the same points throughout the cell (Figure 101 B). The average overlap correlation score for these sections was 0.68 (Figure 101 C). In cells differentiated according to Du *et al* once again a portion of GLuc was sorted into the SVs. There appeared to be little to no effect

149

of the differentiation protocol on the sorting of the VAMP2-GLuc construct, since both protocols result in similar moderate colocalization between GLuc and Syp.
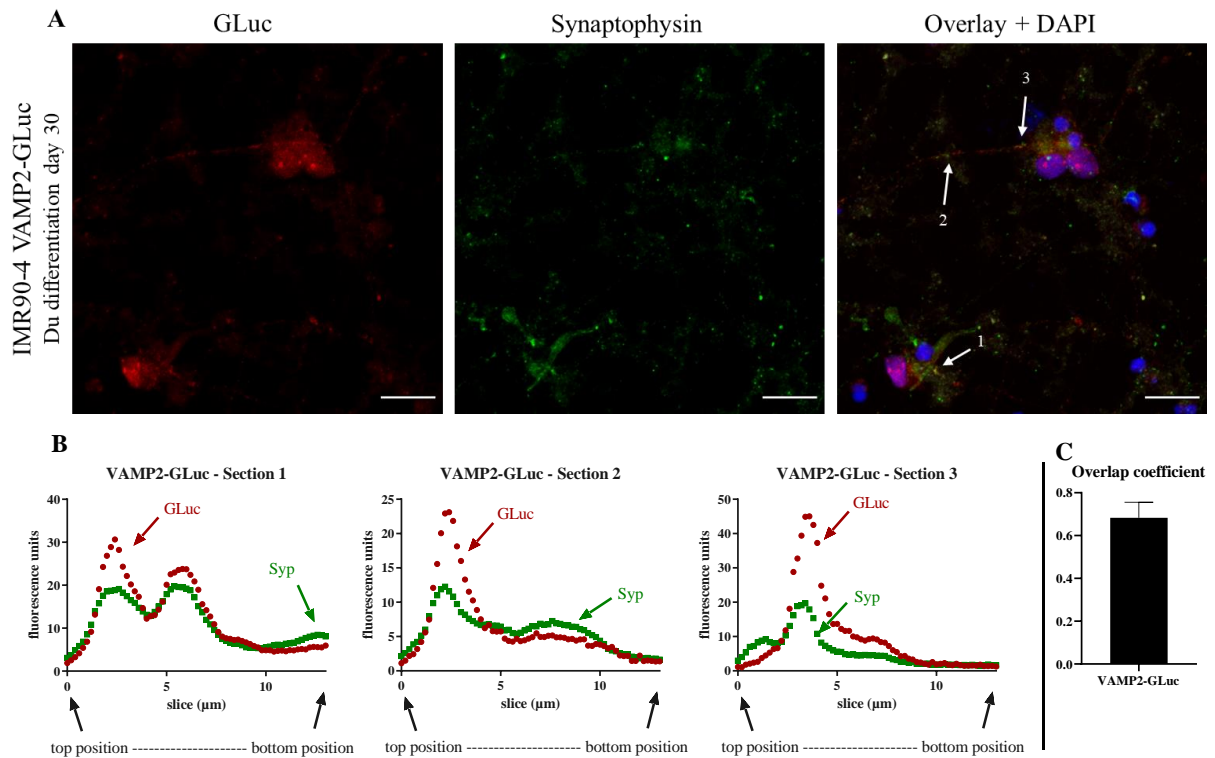
Figure 101: Colocalization analysis of GLuc and Syp in IMR90-4 VAMP2-GLuc MNs (Du D30). Upper panel A: confocal images of immunofluorescence double labeling of IMR90-4 VAMP2-GLuc clone 11. Cells were fixed, permeabilized and then incubated overnight with rabbit-α-GLuc and mouse-α-Syp primary antibodies. Secondary antibody incubation with Alexa-fluor α-rabbit 568 (1:1000) and Alexa-fluor α-mouse 488 (1:1500). The overlay of all images includes DAPI labeled nuclei and sections of colocalization analysis indicated by white arrows in overlay image; scale bar = 20μm. Bottom panel: (B) z-axis fluorescence analysis of sections indicated in section C, tracing the localization of GLuc and Syp. (C) Overlap coefficient for colocalization of GLuc with Syp.

Summary overlap coefficients

There was no statistical difference between the degree of colocalization of GLuc and the Golgi apparatus in the iPSC IMR90-4 hPOMC-GLuc clone and the corresponding degree of colocalization of GLuc and SgII in the cells differentiated according to Maury *et al* from the same clone. There was also no statistical difference between the degree of colocalization of GLuc and the Golgi apparatus in the iPSC IMR90-4 VAMP2-GLuc clone and the corresponding degree of colocalization of GLuc and Syp in the cells differentiated according to Maury *et al* and Du *et al* (Figure 102). Therefore the extent of vesicle-specific sorting in the motor neurons does not differ from the extent of Golgi sorting. Furthermore, in the case of IMR90-4 VAMP2-GLuc, the extent of vesicle-specific sorting in motor neurons does not differ between the two differentiation protocols.
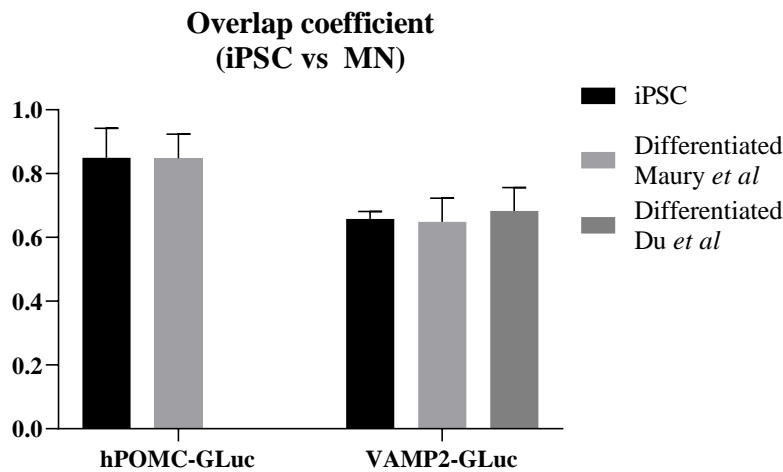
Figure 102: Summary of overlap coefficient in IMR90-4 hPOMC-GLuc and IMR90-4 VAMP2-GLuc iPSCs and MNs. Correlation of GLuc and SgII/Syp for all clones from sections 1-3, in comparison with colocalization of the same clones with the Golgi apparatus analysis with JACoP (Fiji). No significant differences found between iPSCs and corresponding differentiated clone, unpaired t-test, n = 3, mean ± SD.

3.10 Luciferase release in differentiated clones

3.10.1 IMR90-4 hPOMC-GLuc vs IMR90-4 no tag Gluc

The IMR90-4 hPOMC-GLuc and IMR90-4 no tag GLuc clones were differentiated according to Maury *et al* for 32 days. Cells were exposed to Na^+-HBS control buffer and K^+-HBS depolarization buffer with and without 10 mM EGTA for 3 minutes at 37 °C and then luciferase activity was measured in the supernatant. In IMR90-4 hPOMC-GLuc MNs, there was no significant difference in the luciferase activity in the supernatant in the cells exposed to depolarization buffer compared to any of the other buffers. In the IMR90-4 no tag GLuc clone, there was no significant change in luciferase activity released into the supernatant after exposure to control and depolarization buffers (Figure 103). While the luciferase activity detected in the supernatants of the IMR90-4 hPOMC-GLuc clone was significantly higher than the activity detected in the supernatants of the IMR90-4 no tag GLuc clone (f(1)=4.974, p < 0.05), a Sidak post-hoc test did not reveal any significant pairwise differences provoked by exposure to each buffer type.
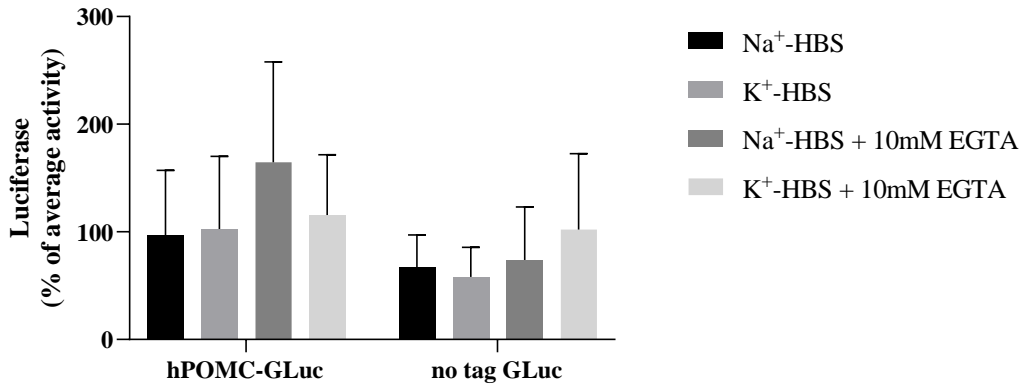
Figure 103: Luciferase released from IMR90-4 hPOMC-GLuc and IMR90-4 no tag GLuc derived motor neurons (Maury D32), 1.5x10⁵ cells were pre-incubated with fresh N2B27 medium for 10 min at 37°C. Release of luciferase activity was stimulated with control (Na+-HBS) or stimulation buffer (K+-HBS) in the absence or presence of 10 mM EGTA for 3 min at 37°C and luciferase activity in the supernatant was measured. The average of the luciferase activity in Na+-HBS and K+-HBS supernatants from each individual clone was calculated and the percent difference from the average for each condition is presented as means ± SD of 24 measurements in 8 independent experiments (hPOMC-GLuc) or 12 measurements in four independent experiments (no tag GLuc). Statistically significant outliers were removed using the interquartile range. Statistical analysis with two-factor ANOVA, post-hoc analysis with Sidak's multiple comparison test.

The IMR90-4 hPOMC-GLuc clone was plated at various confluencies and differentiated according to Maury *et al* for 32 days. Two different pre-incubation media were compared: the standard N2B27 medium and neurobasal medium. Cells were exposed to either standard control or depolarization buffers. Pre-incubation of the samples plated at the standard cell density (1.5 x 10^5 cells/well) with neurobasal buffer resulted in higher luciferase activity from the samples exposed to EGTA (Figure 104 A). With cells plated at a density of 7.5 x 10^4 cells/well, there was a slightly higher luciferase activity in the samples exposed to the depolarization buffers, both without and with EGTA (Figure 104 B). No differences in luciferase activity were evoked by any of the buffers in cells plated at densities between 6 x 10^4 and 3 x 10^4 cells/well (Figure 104 C and D).
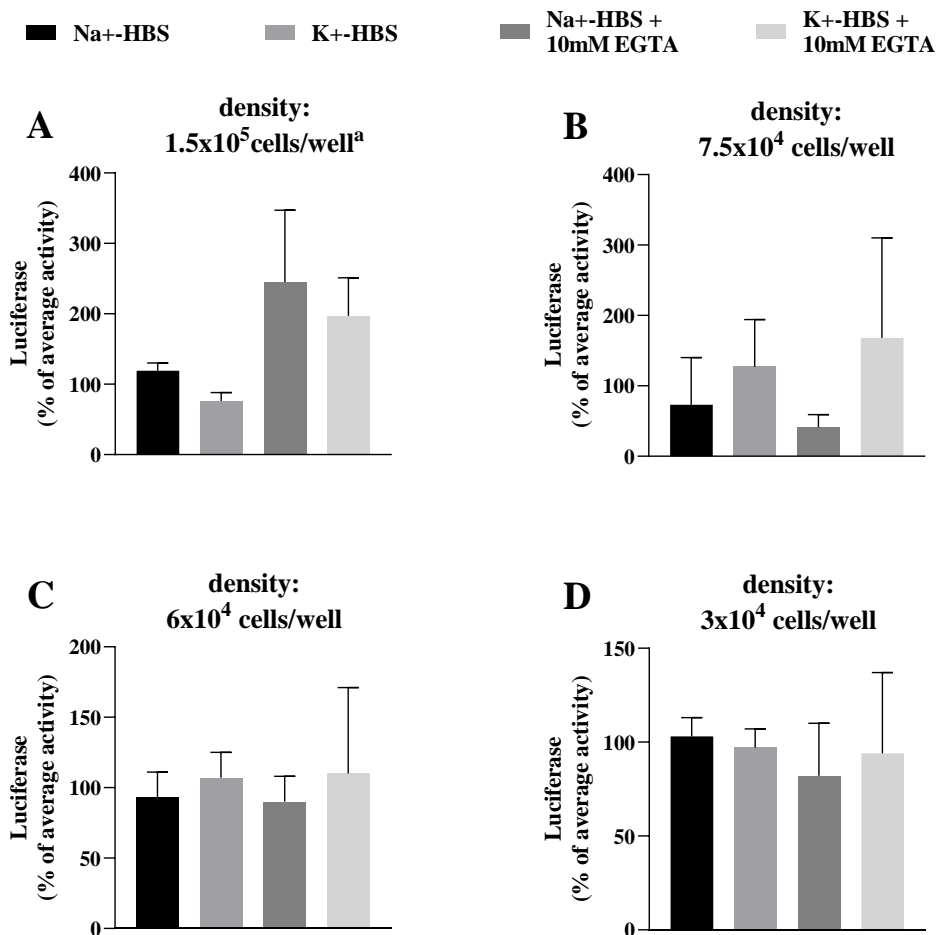
Figure 104: Luciferase released from IMR90-4 hPOMC-GLuc derived motor neurons (Maury D32). Cells were pre-incubated with fresh N2B27 or Neurobasal medium for 10 min at 37°C. Release of luciferase activity was stimulated with control (Na+-HBS) or stimulation buffer (K+-HBS) in the absence or presence of 10 mM EGTA for 3 min at 37°C and luciferase activity in the supernatant was measured. The average of the luciferase activity in Na+-HBS and K+-HBS supernatants from each individual clone was calculated and the percent difference from the average for each condition is presented as means ± SD. Various densities of cells were plated to test the effect of confluency on the test, (a) neurobasal pre-stimulation medium was tested. Sections A, C, D,: n= 4 in 2 independent experiments; B: n = 2 in 1 independent experiment.

3.10.2 IMR90-4 VAMP2-GLuc luciferase release

3.10.2.1 Establishing appropriate experimental parameters

It was first necessary to determine the appropriate TEV protease concentration, incubation time, and temperature needed to detect possible post-depolarization circulating GLuc. The initial experiments including the TEV protease were carried out at 30 °C, due to the enzyme's optimal activity at between 4 - 30 °C[113]. The cells were incubated in either Na$^+$-HBS control buffer and the K$^+$-HBS depolarization buffer in the absence or presence of 5U TEV protease and the absence or presence of the calcium chelator EGTA. There was only a very minor and insignificant trend towards inhibition of luciferase release in those samples incubated

without TEV protease and with 10 mM EGTA. None of the other combination of factors tested resulted in any change in luciferase activity patterns (Figure 105).

**Luciferase release from IMR90-4 VAMP2-GLuc derived motor neurons
20 minute buffer incubation at 30 °C**
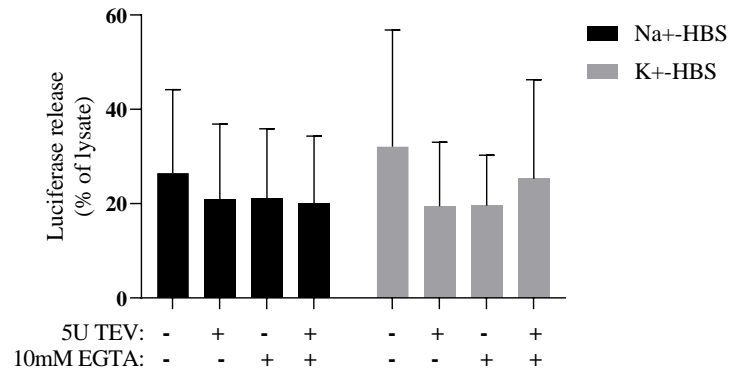
Figure 105: Luciferase release under exposure to 5U TEV protease from IMR90-4 VAMP2-GLuc derived motor neurons(Maury D30). Cells were pre-incubated with fresh N2B27 medium for 10 min at 37 °C. Release of luciferase activity was stimulated by Na$^+$-HBS control buffer and the K$^+$-HBS depolarization buffer in the absence and presence of 5 U TEV protease and 10 mM EGTA at 30 °C for 20 min. The quotient of the luciferase activity in the supernatants and the lysates from each condition was calculated, presented as means ± SD. n = 6 in 2 independent experiments.

Synaptic vesicle endocytosis occurs at a slower rate at room temperature[114], therefore the subsequent experiment was carried out for 2 hours at 24 °C. Cells were incubated with medium, in addition to Na$^+$-HBS control buffer and K$^+$-HBS depolarization buffer, in order to assess if spontaneous GLuc/neurotransmitter release could be measured. A series of increasing concentrations of TEV protease (0 U, 1 U, 5 U, and 10 U) were tested per incubation scenario. There was an insignificant increase in luciferase activity in cells exposed to increasing quantities of TEV protease while incubated with medium. There was no significant increase in luciferase activity upon exposure of the cells to 1 U of TEV protease while incubating with Na$^+$-HBS buffer nor with K$^+$-HBS buffer. There was an insignificant trend towards increased luciferase activity after incubating the cells in Na$^+$-HBS buffer plus 5 U TEV protease, while there was a significant 3 fold increase in luciferase activity after cells were exposed to 5 U TEV protease in the K$^+$-HBS buffer. After incubation with 10 U TEV protease in Na$^+$-HBS buffer the luciferase activity increased by approximately 8 fold and in K$^+$-HBS buffer the luciferase activity also increased by over 7 fold (Figure 106). All subsequent experiments including the TEV protease were carried out with 10 U per reaction for 2 hours at 24 °C.

154

Luciferase release from IMR90-4 VAMP2-GLuc derived motor neurons TEV protease concentration test
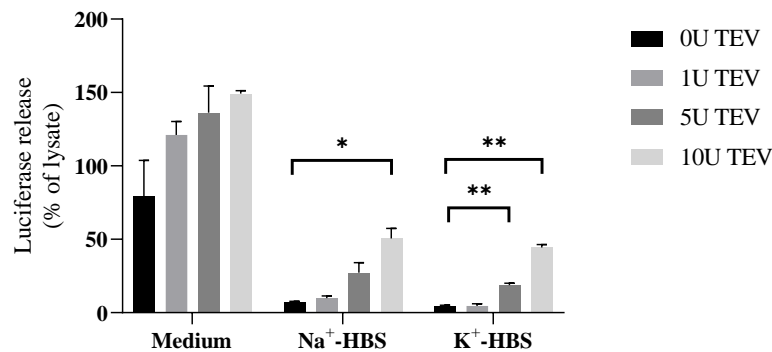
Figure 106: Luciferase release under exposure to increasing concentrations of TEV protease from IMR90-4 VAMP2-GLuc derived motor neurons (Du D31). Cells were pre-incubated with fresh N2B27 medium for 10 min at 37 °C. Release of luciferase activity was stimulated by N2B27 medium, Na+-HBS buffer, and K+-HBS buffer in the presence of 0U, 1U, 5U or 10U TEV protease at 24 °C for 2 hours The quotient of the luciferase activity in the supernatants and the lysates from each condition was calculated, presented as means ± SD. *p < 0.05, **p < 0.01; n = 2 in 1 independent experiment.

3.10.2.2 Luciferase release from IMR90-4 VAMP2-GLuc derived motor neurons exposed to BoNT/A

Due to the consistent increase of luciferase activity in samples incubated in both Na^+-HBS or K^+-HBS buffers and exposed to TEV protease, IMR90-4 VAMP2-GLuc clone 11 was differentiated according to Du *et al* for 38 days and tested for the release of luciferase after exposure to Botulinum neurotoxin serotype A for 48 hours. Two groups of cells (those incubated in standard differentiation medium and those additionally exposed to 100 pM BoNT/A) were incubated in either medium, Na^+-HBS buffer, or K^+-HBS buffer in the absence or presence of 10 U TEV protease. Neither the exposure to TEV protease nor to BoNT/A affected those cells incubated in medium. On the other hand, there was a strong 22 fold increase in luciferase activity in the supernatant of cells incubated in Na^+-HBS buffer while exposed to TEV protease. Exposure to BoNT/A in the parallel cell group significantly decreased the supernatant luciferase activity to only 16 fold higher than the control without the protease. The cells incubated in K^+-HBS buffer and exposed to TEV protease had an 18 fold higher supernatant luciferase activity as opposed to the cells not exposed to TEV protease. While the exposure to BoNT/A appeared to slightly reduce luciferase activity in cells incubated with TEV protease and K^+-HBS buffer, this decrease is not significant (Figure 107).

Luciferase release from IMR90-4 VAMP2-GLuc derived motor neurons incubation ± 100 pM BoNT/A
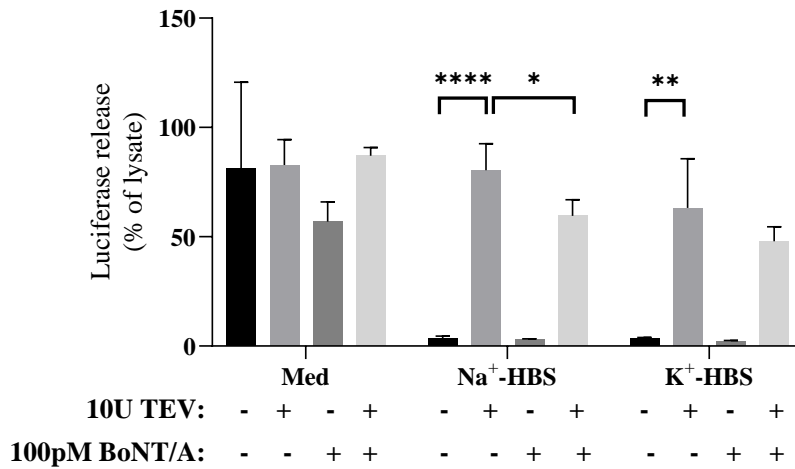
Figure 107: Luciferase release under exposure to TEV protease and BoNT/A from IMR90-4 VAMP2-GLuc derived motor neurons (Du D38). Cells were incubated in either standard differentiation medium or differentiation medium plus 100 pM BoNT for 48 hours. Cells were washed once with fresh medium and then pre-incubated with fresh N2B27 medium for 10 min at 37 °C. Release of luciferase activity was stimulated by N2B27 medium, Na+-HBS buffer, or K+-HBS medium in the absence and presence of 10U TEV protease at 24 °C for 2 hours. The quotient of the luciferase activity in the supernatants and the lysates from each condition was calculated, presented as means ± SD. Statistical analysis determined by t-test. *p < 0.05, **p < 0.01, ****p < 0.0001. n = 4 in 2 independent experiments.

3.10.2.3 Modifications to luciferase release protocol for IMR90-4 VAMP2-GLuc clone

In order to establish a more substantial cell depolarization technique, IMR90-4 VAMP2-GLuc cells differentiated for 38 days according to Du *et al* were incubated in Na$^+$-HBS buffer prepared with and without 10 U TEV protease in the presence and absence of 1 mM carbachol, a muscarinic and nicotinic agonist. As previously determined, the addition of 10 U TEV to Na$^+$-HBS buffer significantly increased the release of luciferase into the supernatant. The addition of 1 mM carbachol to Na$^+$-HBS buffer did not significantly affect luciferase release. In Na$^+$-HBS buffer containing 10 U TEV and 1 mM carbachol, luciferase release increased by approximately 3 fold in comparison to the buffer containing only 10 U TEV (Figure 108). Carbachol in Na$^+$-HBS buffer could therefore instigate luciferase release from IMR90-4 VAMP2-GLuc motor neurons. This is very important because, while exposure to TEV protease alone can release GLuc into the supernatant, which can be partially repressed by BoNT/A with the implication that this degree of reduced GLuc arises from the blocked fusion of SVs to the presynaptic membrane, the increase of GLuc activity in the supernatant after the addition of 1 mM carbachol implicates that a fusion event is being stimulated beyond that measured by TEV protease alone.

156

**TEV-mediated Carbachol-elicited GLuc release
from IMR90-4 VAMP2-GLuc derived motor neurons**
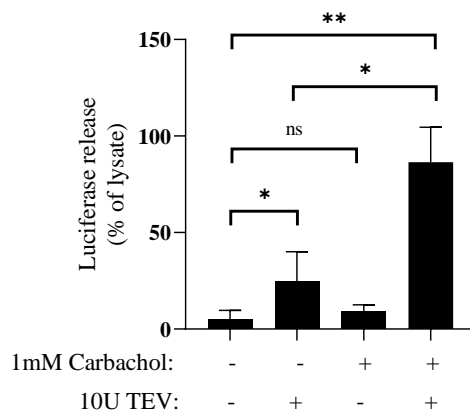
Figure 108: Luciferase release under exposure to carbachol or TEV protease from IMR90-4 VAMP2-GLuc derived motor neurons (Du D38). Cells were pre-incubated with fresh N2B27 medium for 10 min at 37 °C. Release of luciferase activity was stimulated by Na+-HBS buffer including 10U TEV protease in the absence or presence of 1 mM carbachol at 24 °C for 2 hours. The quotient of the luciferase activity in the supernatants and the lysates from each condition was calculated, presented as means ± SD. Statistical significance determined by t-test, ns = not significant, * = p < 0.05, ** = p < 0.01. n = 4-8 in 2-4 independent experiments.

3.10.2.4 IMR90-4 VAMP2-GLuc clone differentiated according to Du *et al*

IMR90-4 VAMP2-GLuc clone 11 was differentiated for 38 days according to Du *et al*. Cells were exposed to N2B27 medium, Na$^+$-HBS buffer and K$^+$-HBS buffer containing 10 U TEV protease in the presence and absence of 1 mM carbachol and 10 mM EGTA. The presence of carbachol induced a significant increase of luciferase release in cells exposed to Na$^+$-HBS buffer and K$^+$-HBS buffer, but not in cells exposed to N2B27 medium. The addition of 10 mM EGTA to Na$^+$-HBS buffer significantly reduced the luciferase released elicited by carbachol by approximately 2 fold. The addition of EGTA also caused a significant 2 fold decrease of luciferase release in cells exposed to N2B27 medium without carbachol, but otherwise no other significant differences in luciferase release (Figure 109).
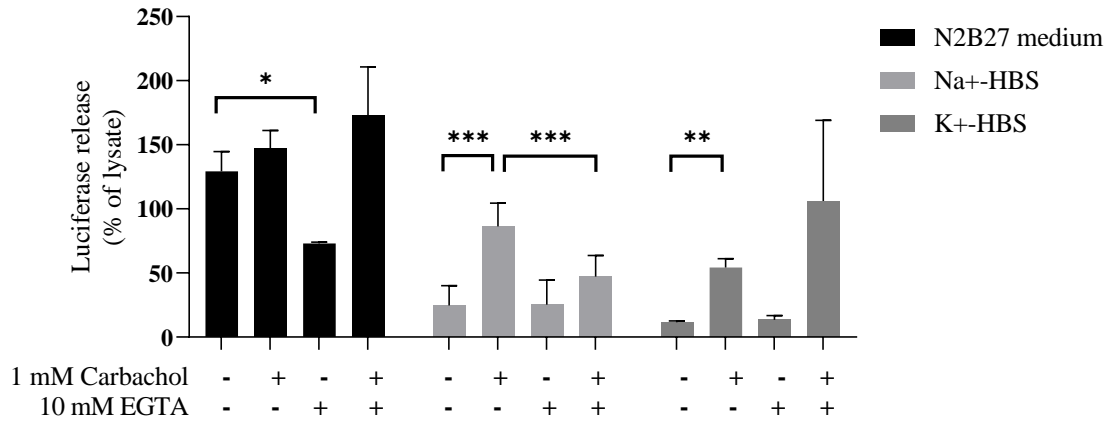
Figure 109: Luciferase release under exposure to TEV protease, and carbachol and/or EGTA from IMR90-4 VAMP2-GLuc derived motor neurons (Du D38). Cells were pre-incubated with fresh N2B27 medium for 10 min at 37 °C. Release of luciferase activity was stimulated by N2B27 medium, Na+-HBS buffer, or K+-HBS medium including 10U TEV protease in the absence or presence of 1 mM carbachol and 10 mM EGTA at 24 °C for 2 hours. The quotient of the luciferase activity in the supernatants and the lysates from each condition was calculated, presented as means ± SD. ** = p <0.01, ***p < 0.001, n = 2-8 in 1-4 independent experiments.

3.10.2.5 IMR90-4 VAMP2-GLuc clone differentiated according to Maury *et al*

IMR90-4 VAMP2-GLuc clone 11 were differentiated for 40 days according to Maury *et al*. Cells were exposed to N2B27 medium, Na$^+$-HBS buffer and K$^+$-HBS buffer containing 10 U TEV protease in the presence and absence of 1 mM carbachol and 10 mM EGTA. The presence of carbachol induced a significant increase of luciferase release in cells exposed to N2B27 medium and Na$^+$-HBS buffer, but not in cells exposed to K$^+$-HBS buffer. The addition of 10 mM EGTA to Na$^+$-HBS buffer significantly reduced the luciferase released elicited by carbachol, reducing luciferase in the supernatant to levels similar to Na$^+$-HBS buffer only. The addition of EGTA caused no other significant differences in luciferase release from cells incubated with N2B27 medium or K$^+$-HBS buffer (Figure 110). The similarity between the EGTA-dependent inhibition of TEV-mediated carbachol elicited GLuc release from iPSC derived IMR90-4VAMP2-GLuc motor neurons differentiated according to Du et al and according to Maury et al indicate that the small, but significant increase of MNs gained by the Du protocol might not make a difference in the clone's ability to release luciferase under the correct stimulation conditions.

**EGTA-dependent inhibition of TEV-mediated Carbachol-
elicited GLuc release from IMR90-4 VAMP2-GLuc derived motor neurons**
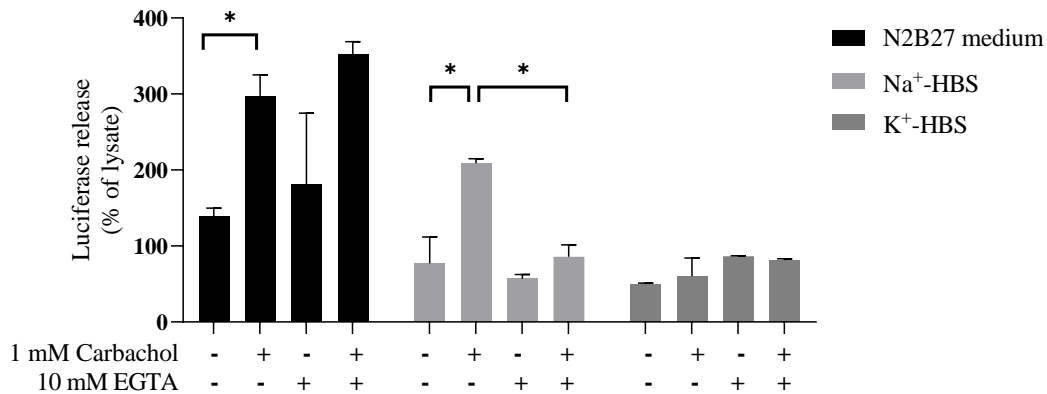
Figure 110: Luciferase release under exposure to TEV protease, and carbachol and/or EGTA from IMR90-4 VAMP2-GLuc derived motor neurons (Maury D40). Cells were pre-incubated with fresh N2B27 medium for 10 min at 37 °C. Release of luciferase activity was stimulated by N2B27 medium, Na+-HBS buffer, or K+-HBS medium including 10U TEV protease in the absence or presence of 1 mM carbachol and 10 mM EGTA at 24 °C for 2 hours. The quotient of the luciferase activity in the supernatants and the lysates from each condition was calculated, presented as means ± SD. *p < 0.05. n = 2 in 1 independent experiment.

3.10.2.6 Specificity of luciferase release from IMR90-4 VAMP-GLuc vs IMR90-4 no tag GLuc derived motor neurons

In order to verify the specificity of the luciferase release from the IMR90-4 VAMP2-GLuc cells in response to carbachol, the IMR90-4 no tag GLuc clone was differentiated for 38 days according to Du *et al* and exposed to Na[+]-HBS buffer in the presence and absence of 1 mM carbachol and 10 mM EGTA. A statistically significant difference in luciferase release was found both between the clone tested (VAMP2-GLuc vs no tag GLuc, f(1)=20.23, p < 0.001) and between the buffers tested (f(3)=6.21, p < 0.01), furthermore there was a statistically significant interaction between the clone and the stimulation buffer (f(3)=10.07, p < 0.001). A Sidak post-hoc test revealed a significant pairwise difference of greater than 6 fold between the luciferase release from IMR90-4 VAMP2-GLuc and IMR90-4 no tag GLuc derived motor neurons provoked by carbachol in Na[+]-HBS buffer (Figure 111).

**TEV-mediated Carbachol-elicited GLuc release from
IMR90-4 VAMP2-GLuc vs IMR90-4 no tag GLuc derived motor neurons**
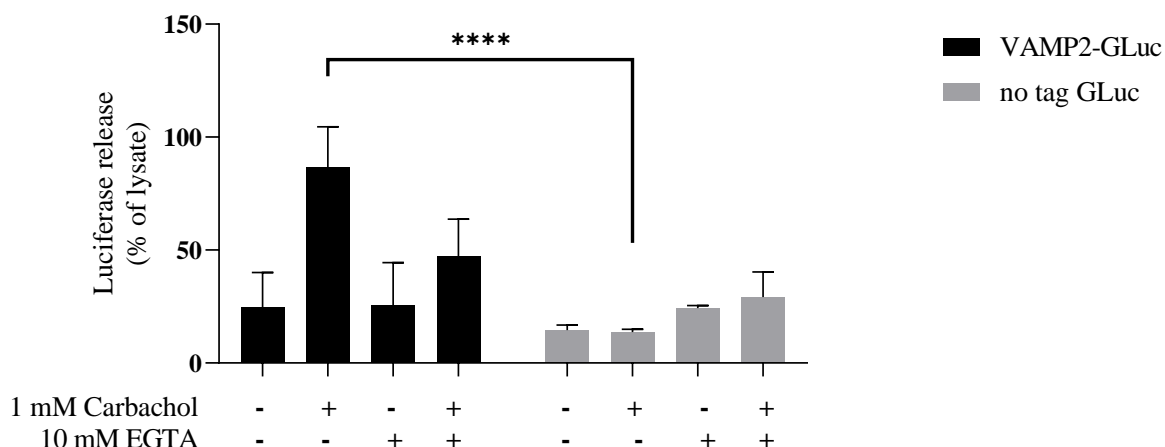
Figure 111: Luciferase release under exposure to TEV protease, carbachol and EGTA from IMR90-4 VAMP2-GLuc and IMR90-4 no tag GLuc derived motor neurons (Du D38). Cells were pre-incubated with fresh N2B27 medium for 10 min at 37 °C. Release of luciferase activity was stimulated by Na+-HBS buffer including 10U TEV protease in the absence or presence of 1 mM carbachol and/or 10 mM EGTA at 24 °C for 2 hours. The quotient of the luciferase activity in the supernatants and the lysates from each condition was calculated, presented as means ± SD. Statistical analysis with two-factor ANOVA, post-hoc analysis with Sidak's multiple comparison test, *** = p < 0.001; n = 2-8 in 1-4 independent experiments.

3.10.2.7 Conclusions of luciferase release optimization

At this timepoint, a protocol to reliably provoke the release of luciferase from the differentiated IMR90-4 hPOMC-GLuc clone has not been developed. On the other hand, the Du *et al* differentiated IMR90-4 VAMP2-GLuc clone showed promising responses to exposure to BoNT/A, whereby the increase in luciferase release in Na$^+$-HBS buffer with 10 U TEV protease was significantly reduced, although not entirely inhibited, after exposure to BoNT/A. Furthermore, when incubated with 10 U TEV protease, IMR90-4 VAMP2-GLuc clone differentiated according to both Du *et al* and Maury *et al*, releases significantly greater amounts of luciferase after exposure to 1 mM carbachol in Na$^+$-HBS buffer as compared to Na$^+$-HBS buffer alone. Furthermore, this luciferase release can be significantly reduced during EGTA and Carbachol co-incubation. This experimental design can be used as a foundation for troubleshooting in the search for an appropriate protocol for the MoN-Light BoNT assay.

4 Discussion

4.1 Overview

4.1.1 Background of motivation behind the MoN-Light BoNT assay development

Botulinum neurotoxin is an important medication to treat multiple disorders associated with hyperactivity of efferent nerves. The prerequisite for the safe use of the potentially lethal toxin is the exact determination of the activity of each batch, which usually is achieved with the ethically controversial mouse lethality assay[43–45]. Alternative methods to accurately and reliably measure the toxin's potency, beyond the standard mouse bioassay, are greatly lacking. Recently, a novel method was developed that allows the determination of the activity of any BoNT serotype by measuring the simultaneous release of neurotransmitters and a genetically engineered reporter protein from SIMA neuroblastoma cells upon cellular depolarization[2]. As discussed in the Introduction (see 1.3.4.3), most other alternatives to animal testing are either only sensitive to one BoNT serotype, or only measure the activity of the LC endopeptidase, without taking into account the important activation steps of the toxin, including the protein's cellular internalization or the translocation of the LC from the SV lumen into the cytosol of the presynaptic terminal (see 1.3.3 for summary of BoNT activity in MNs).

4.1.2 Aim of MoN-Light BoNT assay design

While the potential applicability and relevance of the SIMA prototype assay, carried out with the SIMA Random-Insertion_hPOMC1-26GLuc clone, is unmistakable, there is still room for improvement. The purpose of this project was to identify the aspects of the SIMA Random-Insertion_hPOMC1-26GLuc prototype that could possibly hinder it from being the most accurate and precise assay, and to modify them to mimic the *in vivo* intoxication with BoNT as closely as possible. The vision of this *in vivo* simulation entails the use of motor neurons, the target cells of BoNT, which have been differentiated from hiPSCs modified with the CRISPR/Cas9 genetic engineering tool to stably express a Gaussia luciferase reporter gene. CRISPR/Cas9 allows the insertion of the donor DNA into an exact point in the genome, in this case the AAVS1 safe harbor locus, to avoid any of the unwanted side effects potentially caused by other modification methods which randomly modify the genome[69,70]. The genetic engineering of the hiPSCs allows the unlimited propagation of human-derived cells which can be differentiated into any cell type while still carrying the donor DNA integrated into the cell's genome. The differentiation into motor neurons, which should express the exact receptors

necessary for the uptake of BoNT and the exact target proteins cleaved by the assorted BoNT serotypes, creates a hypothetically ideal environment in which BoNT can function. The implementation of these three factors together can greatly increase the applicability and relevance of this potency assay.

4.1.3 Characterization and problematic aspects of clones prepared for the MoN-Light BoNT assay

In order to assure that the clones developed for use in the MoN-Light BoNT assay were accurately designed, that the CRISPR/Cas9 method was correctly implemented, and that the resulting clones were expressing the reporter protein in the correct subcellular compartments, the clones were fully characterized. Post-transfection and isolation of the clones, it was verified that they contained the reporter gene in the AAVS1 safe harbor locus, the number of insertions of the reporter gene in the genome was analyzed, and the localization of the reporter protein was identified. Only after all aspects of the genetic modification were analyzed could the clones be tested for their applicability in the MoN-Light BoNT assay. Once it became clear that the luciferase release conditions optimized for SIMA Random-Insertion_hPOMC1-26GLuc could not be transferred and applied directly to the differentiated MNs derived from the IMR90-4 hPOMC-GLuc clone, the additional requirement arose to address which aspects of the clone preparation could be problematic. Several potential issues were identified:

1) The movement of GLuc through the neurosecretory pathway was partially verified by the protein's colocalization with the Golgi apparatus. However, it was still unknown whether upon differentiation, the hPOMC signal peptide continued to correctly direct GLuc into neurosecretory vesicles. To troubleshoot this problem, two additional constructs to direct GLuc into LDCVs with the CgA and SgII sorting tags were prepared.

2) The hPOMC signal peptide directs GLuc into LDCVs, which are similar to SVs, from which classical neurotransmitters are released, but which are distinctly packaged and undergo divergent exocytosis (see 1.2.1 in Introduction). The assay ideally should measure the exact co-release of neurotransmitters and GLuc from the same vesicle, so a construct was designed to incorporate GLuc into secretory vesicles by fusing GLuc to the transmembrane SV protein, VAMP2.

3) While the expression of GLuc was verified in the MN population tested in the release assay, the population is actually not 100 % MNs[99]. It was possible that GLuc was not

162

actually being expressed in the correct cell subpopulation. In order to exclude this possibility, it would be necessary to verify the colocalization of GLuc with proteins associated with differentiated motor neurons.

4) The fusion of LDCVs and SVs to the presynaptic membrane of neurons is likely instigated by different factors, whereby a brief stimulus can provoke extremely quick exocytosis of SVs, and a stronger stimulus can provoke a delayed but longer-lasting exocytosis of LDCVs[114–116]. The depolarization protocol was optimized in SIMA cells, which might respond differently to stimuli than the iPSC-derived MN population. Therefore it is necessary to optimize a specific protocol for the differentiated cells.

4.2   Production of CRISPR/Cas9 associated plasmids

The CRISPR/Cas9 genetic engineering in this project was carried out by the co-transfection of two plasmids. One plasmid contained the coding sequence for eSpCas9, an enhanced specificity Cas9 developed by the Zhang lab[72] designed to reduce off-target effects of the CRISPR/Cas9 system. This plasmid also contained a cloning site in which the sgRNA sequence should be integrated, which directs the Cas9 endonuclease to the unique cleavage site in the genome. The AAVS1 genomic safe harbor locus, located in the *PPP1R12C* gene, was selected due to its wide use and the availability of multiple validated plasmids for the site. There are two common target sites at the AAVS1 safe harbor locus, designated as T1 and T2, to which the sgRNA can direct the endonuclease. An analysis of successful homology-directed recombination integrations at these two sites found there was an approximate 3 % success rate at site T1, while there was an approximate 8 % success rate at site T2[73]. For this reason, the sgRNA recognizing the T2 site was cloned into the eSpCas9(1.1) plasmid (Figure 23).

The second plasmid, pAAVS1-P-MCS, was already designed to contain the sequences for the left and right homology arms for the AAVS1 safe harbor locus and the puromycin antibiotic resistance gene[74]. The remaining donor DNA sequence between the two HAs, including the promoter and Gaussia luciferase fused with a selection of sorting signals, were cloned into the MCS. Two derivatives of the Ef1α promoter were initially prepared and incorporated into the donor plasmid. The Ef1α promoter has been reported to maintain consistent and strong expression in both pluripotent and differentiated states[107,109,117] and was deemed to be an appropriate promoter for this project. The Ef1α promoter was amplified from gDNA extracted from HEK cells using a primer pair designed according to the promoter sequence information provided with the plasmid pEF-Bos[118]. The second promoter, designed

by InvivoGen, is a hybrid of a segment of the EF1α core promoter fused with a segment of the Human T-Cell Leukemia Virus (HTLV). The Ef1-HTLV promoter is non-tissue specific and highly expressed in all cell types, the addition of the HTLV segment should increase steady state transcription and translation efficiency by enhancing RNA stabilization[97]. Both the promoters were successfully cloned into the pAAVS1-P-MCS donor plasmid, along with the hPOMC-GLuc fusion sequence (see section 3.1.2). SIMA neuroblastoma cells were transiently transfected in parallel with both constructs, however, only the Ef1-HTLV fusion promoter actually led to the expression of GLuc (Figure 32). The explanation why the Ef1α promoter did not express GLuc in the SIMA cells remains unclear. No troubleshooting was carried out for this promoter, since the second promoter option, Ef1-HTLV, highly expressed GLuc.

The first sorting signal used to direct GLuc into secretory vesicles was the same first 26 amino acids of hPOMC used for SIMA Random-Insertion_hPOMC1-26GLuc. This seemed to be the obvious choice, since it was already shown by differential centrifugation that GLuc was directed into vesicles and furthermore the clone was already proven to reliably release luciferase upon exposure to depolarization buffer[2]. To account for inefficient sorting or targeting into an incorrect vesicle population, several additional plasmid constructs to sort GLuc to various vesicle types were prepared. These alternative sorting signals included CgA, SgII, and VAMP2. The granin family, including CgA and SgII, are well-characterized proteins in the regulated secretory pathway[20] and should therefore sort GLuc into the same LDCVs as hPOMC. On the other hand, VAMP2 is a transmembrane protein on both LDCVs and SVs and is a member of the SNARE protein family subject to cleavage by BoNT/B, D, F, and G[24]. In this construct, GLuc was fused to VAMP2 by a linker sequence containing multiple restriction enzyme recognition sites and a recognition sequence for the TEV protease[119]. Both the TEV recognition sequence and GLuc were designed to extend into the lumen of the vesicle, attached to the C-terminus of VAMP2. This construct should increase the possibility that GLuc actually reaches the SVs, which contain classical neurotransmitters and are therefore the most relevant vesicles to neuronal communication and the paralytic effects of BoNT[82,84,116]. The use of the TEV recognition sequence linker between VAMP2 and GLuc was incorporated in order to provide a possibility to selectively cleave GLuc from VAMP2 at the point of vesicle-presynaptic membrane fusion (see section 1.6.1.3 for a full description of GLuc/VAMP2 cleavage).

4.3 Characterization of GLuc clones

Once clones were produced and selected for puromycin antibiotic resistance, it was important to characterize them, as mentioned above. First, it was necessary to verify that the insertion of the donor DNA took place at the AAVS1 safe harbor locus, as directed by CRISPR/Cas9. Second, the specificity of the integration must also be verified to track any off-target donor DNA integrations. Finally, it was necessary to confirm that the GLuc protein is sorted into the targeted neurosecretory vesicle.

4.3.1 Confirmation of donor DNA integration at the AAVS1 safe harbor locus

The first two steps are necessary to verify the efficacy of the genetic modification using CRISPR/Cas9. The genetic editing tool can be applied to insert a donor DNA into a specific locus in the gDNA by homology-directed recombination, and therefore the correct integration with no off-target insertions must be verified. The first step in each clone's characterization is theoretically easy. The target locus is known, therefore the gDNA sequence is known, and PCR primers can be designed to flank the homology arms surrounding the locus. These primers will amplify either the unaffected WT sequence or, if the donor DNA is successfully inserted, the modified sequence. The theory is straightforward, but in this case it required substantial optimization (sections 3.2.1 and 6.5). The first attempts to amplify the AAVS1 safe harbor locus in gDNA from non-transfected SIMA cells using primer pairs located 5' of the left HA and 3' of the right HA were unsuccessful despite testing multiple primer pairs at multiple annealing temperatures. Furthermore, the control amplification with the same primers and thermocycling protocol published by the group which produced and distributed the donor DNA plasmid[74] also did not result in an amplified product. Finally, using a primer localized in the middle of the right HA, paired with the original primer located 5' of the left HA, a product of the correct size with no non-specific by-products could be amplified from gDNA of non-transfected cells. This same primer pair could also be used to amplify the locus containing the integrated donor DNA, thus identifying both heterozygous and homozygous integrations.

The use of guanine/cytosine (GC)-rich genomic templates can have an adverse effect on PCR amplification efficiency, whereby GC residues can fold into complex secondary structures which do not melt and remain inaccessible to PCR primers[120]. The 200 bp flanking the 3' side of the right HA contain 64 % GC content, while the final successfully amplified insert confirmation region contains 57 % GC content. Despite PCR protocol optimization with DMSO, the high GC content could have hampered the amplification reaction. While is not ideal

that one of the primers is not completely outside the area effected by the HDR, the sequencing results of the homozygote clones have confirmed that the HA regions retain the exact expected sequence, therefore the slightly shorter coverage of the insert confirmation product does not affect the identification of donor DNA integration at the AAVS1 safe harbor locus (see Figure 14, Figure 42, section 3.2.1, and section 6.4.10).

As documented in the Results (section 3.4.1), of all the isolated CRISPR-modified SIMA clones, only one (SIMA hPOMC-GLuc 29) appeared to have gone through full homozygous donor DNA integration. All remaining clones with a confirmed insert were heterozygotes for the donor DNA and WT alleles. The rate of homozygosity appears to be 1 out of 21 clones, and therefore less than 5 %. On the other hand, of the total 77 CRISPR-modified IMR90-4 clones analyzed for their insert status, 50 of them were homozygote clones, reaching a rate of almost 65 %. The main mechanisms of double strand break repair are homology-directed recombination and non-homology end joining. HDR is highly specific and mainly active during the S/G$_2$ (synthesis/gap 2) phases, whereas NHEJ is more prone to errors but constitutively active throughout the cell cycle[121]. Immortal tumor-derived cell lines, such as the SIMA neuroblastoma cell line, are inherently flawed in that they have chromosomal abnormalities which allow them to continuously divide. In addition they suffer from defective DNA repair mechanisms[122]. Therefore, it might be the case that the more common and less precise DNA repair mechanism, NHEJ, is favored by the SIMA cells, decreasing the chances for HDR to take place and leading to the very low rate of homozygosity in comparison to the IMR90-4 cells.

The clones that would be potentially used in the final differentiation and MoN-Light BoNT assay were sequenced by Sanger. This was done to identify any possible mutations and to verify that the insert amplicon visualized as a band in the agarose gel really represented the donor DNA inserted in the AAVS1 safe harbor locus (see sections 3.4.1.1 and 6.4). The sequence confirmation was carried out using the primary insert confirmation amplicon and an additional secondary amplicon covering the right homology arm. As previously discussed, both of the primary insert confirmation primers do not flank the entire sequence of both the HAs. Rather, one primer is located in the middle of the right HA, while the other is found outside of the left HA. An extra secondary amplicon covering the entire right HA and part of the Ef1-HTLV promoter was designed to be complementary to the primary insert confirmation product, in order to ensure 100 % coverage of the most important areas around the AAVS1 safe harbor

166

locus: both HAs, the promoter sequence, and the tagged GLuc sequence. The sequences of the hPOMC-GLuc clones 6 and 4 clones aligned perfectly to the reference genomic sequence except for one base, a C>T mutation that was found in both clones at the same nucleotide. This position was found in the UCSC Genome Browser and identified as the SNP rs667451[123], the T variant is in fact the major allele with a frequency of 91.6 %, making it unsurprising that either the IMR90-4 cells were homozygote T-allele carriers for this SNP, or that the sequence from the 3' HA of the donor plasmid was incorporated into the genomic DNA at this point during HDR. The unequivocal output of the sequences surrounding the HAs is a testament to the efficacy of HDR, since this genomic area suffered a double strand break on both alleles and completely integrated foreign DNA using a plasmid template without a single base change on either allele.

4.3.2   Analysis of CRISPR-associated off-target integration events

An important aspect in the development of the MoN-Light assay is the exclusion of clones with random integrations of the donor DNA into their gDNA. This is to prevent any potential disruption in normal cellular processes that could be caused by the integration of foreign DNA into functionally relevant genes. Therefore, the second essential element in the characterization of the CRISPR-modified clones is to provide evidence of which clones underwent multiple off-target insertions and which clones were successfully modified with only 1 - 2 copies of the donor DNA in the AAVS1 safe harbor locus. Three techniques were tested to this end, Southern blot, ligation-mediated PCR, and double-control qcnPCR.

4.3.2.1   Southern blot

Southern blot is a standard method to detect copy number variations in genomic DNA[89]. While the southern blot probes were capable of detecting extremely high copy numbers of the donor DNA, in this project it was not proven to be a reliable method to detect single insertion events. As shown in the Appendix 6.6.1, the blots hybridized using the probe GLucF2R2 (Figure 131) could be developed to identify very faint bands at the fragment size known to contain the GLuc donor DNA. However, with increasing exposure times these bands never became more easily identifiable, rather the background increased in corresponding intensity. Furthermore, not a single GLuc probe tested hybridized with any segment of digested gDNA from the SIMA Random_Insertion-hPOMC1-26GLuc clone. This clone also has GLuc incorporated into its genome at multiple unknown loci and therefore should have had multiple

identifiable bands. The method was not sensitive and reliable enough to exclude potential off target integrations of the construct.

4.3.2.2 Ligation-mediated PCR

Another common method to identify insertions of a known DNA fragment in an unknown gDNA locus is the ligation-mediated PCR[124]. This method involves digesting gDNA with a restriction enzyme, then ligating annealed oligonucleotides with the corresponding restriction digested overhangs, and finally PCR amplifying fragments containing the donor DNA. The PCR is carried out using a primer based on the known donor DNA sequence and one of the corresponding ligation oligonucleotides as the primer pair. Unfortunately this method also failed. Since potential qPCR techniques to quantify the insert copy number were being optimized in parallel, the focus was shifted to these tests and away from the ligation-mediated PCR.

4.3.2.3 Double-control quantitative copy number PCR

Introduction to dc-qcnPCR method

Given the fact that qPCR can be used to detect small changes in genetic expression by monitoring the amplification of cDNA fragments[125], it is logical that the same instrument should be capable of also detecting small variations in gDNA copy number. Therefore, an autosomal/ChrX double-control quantitative copy number PCR (dc-qcnPCR) method was devised and optimized to reliably detect one copy (heterozygous insertion), two copies (homozygous insertion, when verified by insertion confirmation gel), or more than two copies (multiple off-target insertions) of a target sequence. Critical to this test is that the cell population is monoclonal. While any result indicating more than two copies clearly suggests the presence of off-target integrations, any result indicating two copies or less could either suggest the heterozygous or homozygous insertions, but could also be due to contaminations from non-GLuc expressing cells. All clones in this project were subjected to monoclonal isolation (see section 2.4.2). The clones used in this project were confirmed to be monoclonal by the detection of GLuc using immunofluorescence (see section 3.4.3)

This technique has proven to be an efficient and simple method to identify clones with only one or two copies of the donor DNA inserted into the genome. If two copies are amplified, it is possible to determine if the clone is homozygous for the insertion or if it contains one on-

target insertion and one off-target insertion in conjunction with the insert confirmation PCR result. The clones isolated for this project were validated through the comparison of GLuc amplification with that of a ChrX gene, after normalization to an autosomal gene. It may be possible to accurately quantify copy numbers greater than two, but this was not validated for this specific project because any copy number greater than two automatically revealed the existence of off-target integrations. In order to validate copy numbers greater than two, it would be necessary to include the sequence of a copy number variant with multiple known copies in the experimental parameters. For the case of this project, however, the most important point was to exclude off-target integrations, therefore the level of sensitivity of this technique was sufficient.

Analysis of CRISPR modified clones

While a selection of the SIMA clones modified to contain the tagged-GLuc construct were shown to contain only two copies of GLuc (for example SIMA hPOMC-GLuc clones 4 and 5), many SIMA-derived clones from both the hPOMC-GLuc group and the VAMP2-GLuc group had off-target integrations in their genomes, sometimes more than ten times as many as expected and desired (see section 3.4.2). This is especially interesting upon consideration that the SIMA clones were also those in which 95 % of the integrations were only heterozygous. One allele would have been subject to an integration event, but the other allele would not take up the donor DNA, while apparently many other disruptions were occurring throughout the gDNA, allowing the integration of the donor DNA in multiple random locations. On the other hand, there was no instance in the CRISPR-modified IMR90-4 clones in which the copy number was more than 4 fold higher than expected. It has been previously shown that off-target effects are common in cell lines with dysregulated repair pathways, such as the homology-driven recombination exploited in this project; whereas off-target effects are rare in healthy human iPSC clones with functional repair pathways[126,127]. Therefore, as similarly discussed above in regard to the difference in homozygosity between the SIMA derived clones and the IMR90-4 derived clones, the abnormalities associated with the immortalized SIMA cell line appear to have disrupted the precise gene editing expected from CRISPR technology. Despite the intact repair pathways present in the IMR90-4 cell line, there were still some off-target events identified, suggesting that the direction of the Cas9 endonuclease to the AAVS1 safe harbor locus was not always perfect, but still more precise in the iPSCs than in the SIMA cells.

Off-target activity of CRISPR/Cas9 technology

Despite the great hype and excitement surrounding the accuracy of the CRISPR/Cas9 method, in the years since its establishment much work has been invested in producing delivery systems with increased efficiency and precision[72,128–130]. One unfavorable point of the technology is that the DSB induced by the Cas9 endonuclease is modulated by a short 20 nucleotide sgRNA, which upon direction to cleavage sites in the genome is actually tolerant to mismatches in the target DNA[130]. This lack of specificity in the identification and targeting of genomic loci increases the risk of off-target DSBs and therefore also of off-target donor DNA integrations. While there is little room to modify the short sgRNA, it has been shown that specific sgRNA nucleotides distal and proximal to the PAM are correlated with on-target efficiency. The GC content and secondary structure of the sgRNA also influence on-target efficiency[131], points which might be exploited to improve editing specificity.

On the other hand, much more focus has been placed on modifications of the Cas9 endonuclease to enhance on-target DSBs. These approaches include either using Cas9 nickase mutants to nick a single-strand of gDNA on either side of the insertion locus, or reducing the expression of Cas9, or identifying and modifying specific amino acids which interact with the target DNA and can influence the stability of on and off-target DSBs[72,132,133]. Some of these strategies include aspects that are potentially detrimental to the method: eliciting a DSB with Cas9 nickase mutants involves designing two sgRNAs, while the reduction of active Cas9 during the transfection may decrease on-target modifications as well as off-target modifications. The application of "rationally engineered" Cas9 nucleases[72] seems to be a logical and simple approach to improve on-target efficiencies, but even the enhanced specificity Cas9 endonuclease expressed from the plasmid used in this project (section 3.1.1) did not generate 100 % on-target integrations. This particular project highlights and reinforces the absolute necessity to inspect and validate each CRISPR-modified cell line. With the characterization techniques presented in this work, however, it was possible to isolate and identify homozygote clones with no off-target donor DNA integrations, which can be implemented in the MoN-Light BoNT assay.

4.3.3 Cellular localization of Gaussia Luciferase

In addition to the establishment of the location and number of donor DNA integrations in each clone's genome, the characterization of the CRISPR-modified clones also included the verification of the GLuc protein localization. In all of the clones, except those originating from

170

the no tag GLuc construct, a signal sequence was fused to GLuc, which directs the protein to be trafficked through the Golgi apparatus and packaged in neurosecretory vesicles.

4.3.3.1   Differential centrifugation

The theoretical association of GLuc with the Golgi apparatus was first investigated by differential fractionation. A similar method had been used to characterize the SIMA Random_Insertion-hPOMC1-26GLuc clone, successfully establishing that the subcellular distribution of GLuc in no tag vs hPOMC tagged Random Insertion clones had an inverse relationship. Specifically, after homogenization, 80 % of GLuc in the SIMA Random Insertion no tag clones was found to be in the soluble fraction, while only 20 % of GLuc was found in the vesicular fraction. The exact inverse was measured for the SIMA Random_Insertion-hPOMC1-26GLuc clone, supporting the theory that the hPOMC sorting signal was correctly directing GLuc into secretory vesicles[2]. A similar experiment was carried out using the IMR90-4 hPOMC-GLuc clone 6, homogenizing the cells according to the SIMA protocol and centrifuging the cells at increasing velocities to separate them into 3 fractions: cellular debris, cytosolic, and membrane/vesicle-associated. Unexpectedly, the greatest luciferase activity was clearly detected in the cellular debris and cytosolic fractions (Figure 140). The most likely cause of this discrepancy is that the optimal homogenization protocol is distinct for the SIMAs and iPSCs, therefore the implementation of the SIMA homogenization protocol in IMR90-4 cells resulted in inappropriately disrupted cellular structures and the detection of GLuc in incorrect subcellular fractions.

The successful division of the subcellular components depends greatly on the initial homogenization of the cells. It is necessary to find an intensity to homogenate the cells which will break the cellular membrane, but not destroy the inner components of the cells. Tissue culture cells are considered to be more difficult to fractionate than animal-derived tissue, due to probable differences in cytoskeletal organization. The parameters that influence the quality of cultured cell homogenization include the growth condition of the cells of interest, the homogenization buffer, and the device used for the homogenization[91]. The ideal homogenate is composed of intact organelles and cellular components as distinct elements in a free suspension. Pre-existing cellular organization can cause certain organelles to remain associated with cytoskeletal elements or remain in aggregates. The aggregates are often associated with DNA spillage from ruptured nuclei that can easily sediment. The homogenization process must be optimized for each cell culture line, due to their individual cytoplasmic and cytoskeletal

organization. Part of the ideal optimization necessary to establish an appropriate homogenization protocol involves the analysis of the homogenate with phase contrast microscopy to verify that the nuclei remain intact and that no aggregates have formed[91,92].

The protocol used for this experiment was optimized for SIMA cells[2], but the cells that were analyzed in this case were the IMR90-4 hPOMC-GLuc iPSCs. Since the appropriate control and tools to optimize and verify the homogenization were missing, this experiment was not pursued further. While the use of subcellular fractionation is a classical biochemical method to analyze neurosecretory vesicle localization, it has also been long established that the use of indirect immunofluorescence in combination with confocal microscopy is a relevant method to analyze intracellular sorting and the secretory pathway of both synaptic vesicles and large dense core vesicles[134].

4.3.3.2   Localization of signal peptide tagged GLuc by immunofluorescence

Antibody verification and optimization

The second technique applied to verify GLuc sorting is immunofluorescence. This method is more powerful than differential centrifugation because specific proteins can be detected with antibodies and labeled with fluorescent markers, not only enabling the visualization of the protein of interest, but also the comparison of two different proteins within the same cell in live cells or at a fixed timepoint[135,136]. As described in the Results (section 3.4.3), the protein of interest can be co-stained with a protein which is associated with a particular cellular region or structure and the colocalization between the two proteins can be analyzed[93,94]. All antibodies used for the colocalization analysis were targeting proteins which are known to be associated with the Golgi apparatus, LDCVs, or SVs.

The Golgi was labeled in iPSCs because it is a key organelle in the processing and packaging of proteins that are destined for secretory vesicles[17,110]. If a protein which has a signal peptide for vesicular packaging is associated with the Golgi, it is very likely because the protein is about to be loaded into/on a vesicle. For this reason, the undifferentiated IMR90-4 CRISPR-modified clones were all analyzed for colocalization with the Golgi apparatus (section 3.4.3). The IMR90-4 CRISPR-modified clones were not analyzed for colocalization with any vesicular markers because the iPSCs do not express these proteins at this point. On the other hand, the SIMA CRISPR-modified clones as well as the MNs derived from the IMR90-4 CRISPR-modified clones could be analyzed for the association of GLuc with the vesicular proteins

172

(section 3.9.3). The hPOMC-GLuc constructs were co-stained using antibodies targeting CgA and SgII, because all three of these proteins contain signal peptides for the LDCVs[20,82,137]. The VAMP2-GLuc constructs were co-stained with an antibody targeting Syp, the typical SV marker[79,83,84].

The optimization of the staining protocols and finding the suitable antibody to specifically label the protein of interest can sometimes be arduous. Problematic points faced in this project were the validation of an appropriate Golgi marker and the screening of the anti-Islet1 antibody 1B1. An antibody against the *trans*-Golgi network (anti-Golgin-97) and a Lectin *Helix pomatia* agglutinin Golgi marker were tested before successfully and specifically labeling the *cis*-Golgi network with an anti-GM130 monoclonal antibody (see sections 3.2.3.2 and 6.6.3.2).

Golgin-97 is an important member of the *trans*-Golgi network conserved domain, GRIP (**G**olgin-97, **R**anBP2α, **I**mh1p and trans golgi **p**230), which functions to maintain the integrity of the TGN through its association with TGN resident proteins[138,139]. The specific location of the epitope against which the anti-Golgin-97 clone CDF4 reacts is not elaborated in the product information[140]. This antibody has been successfully used to identify the TGN by other groups in various cell types[141–143], therefore, the lack of staining in this particular sample could be due to the lot production, although no problems had been reported at the time with this production lot. Furthermore, the problem could have been due to the paraformaldehyde fixation of the sample, during which it is possible that the epitope that should be recognized by the antibody was modified, resulting in epitope masking[135,144,145]. Finally, the affinity of the antibody to its epitope might be weak[135,136]. Since it was possible to eventually label the *cis*-Golgi network with the monoclonal anti-GM130 antibody using the standard IF procedure, it is evident that the entire Golgi apparatus structure was not destroyed by any step of the procedure. Rather, the most likely complication was the availability of the epitope to the antibody in combination with lower than optimal affinity of the antibody to the epitope. The unsuccessful binding of the Lectin HPA marker to the α-N-acetylgalactosamine residues associated with the Golgi apparatus was more likely due to the glycosylation state of the organelle, which may have limited the reactivity of the lectin to the *cis*-Golgi cisternae[146,147].

The two antibodies tested in order to label Islet1 were both monoclonal antibodies designed to target amino acids 149-350 of ISL1, although the target epitope of each antibody is not specified by the producer[148,149]. The anti-Islet1 antibody from clone 1B1 was found to associate with many unspecific cell-types, whereas the anti-Islet1 antibody from clone 1H9 was

found to only stain selected cells within a putative motor neuron population (see Figure 55 and Figure 148). The region containing amino acids 149-350 encodes two common families and domains, a LIM binding domain and a glutamine rich section of protein, both of which might contain sequence similarities and therefore epitope similarity between distinct targets throughout the genome[150]. The LIM binding domain found in Islet1, for example, has 80 % similarity to Protein tyrosine phosphatase receptor type C-associated protein, a widely expressed transmembrane protein[151] and 88 % similarity to the Transmembrane protein 231[152]. If the monoclonal Islet1 antibody from clone 1B1 happens to target an epitope that is associated with one of these common regions, then it is possible that the antibody will stain multiple proteins non-specifically. On the other hand, since the monoclonal Islet1 antibody from clone 1H9 does appear to specifically stain motor neurons, it could be the case that the epitope targeted by this antibody is associated with a region unique to Islet1.

Furthermore, discrepancies in labeling capacity between cell lines also arose. For example two different antibodies targeting CgA were tested. Both intensely labelled CgA in SIMA cells, but had no reactivity in MNs (3.2.3.2). It has been shown in multiple studies, however, that CgA is expressed in MNs[153–155], so it is unclear in this case if these MNs actually do not express CgA or if there is another factor, such as epitope masking discussed above, obstructing the protein's detection. The immunohistochemistry protocol and access of the antibody to the LDCV were validated by the positive detection of SgII in MNs. Both CgA and SgII are found in the same set of LDCVs and therefore the access of the antibodies to these proteins would be expected to be similar using the same protocol.

Description of colocalization analysis by confocal microscopy

One of the most important applications of fluorescence microscopy is the comparison of the distribution of two fluorescently labeled molecules in live or fixed cells[94]. Often of interest is whether these molecules are colocalized, which means they are present in proximal spatial regions in a sample[95,156]. Fluorescence colocalization is especially relevant to assess whether a particular molecule or protein associates with a specific cellular structure. The repeated coincidence of two fluorescently labeled molecules can indicate that these two probes are indeed spatially correlated, without the need for high resolution fluorescence resonance energy transfer or electron microscopy[94]. The confocal microscope is capable of optically sectioning the specimen of interest, and therefore can capture images in the z dimension, thereby visualizing potential colocalization in a multidimensional array[95].

174

Colocalization is often subjectively judged by superimposing two images capturing probes labeled with Alexa-488 (green) and Alexa-568 (red) fluorescent secondary antibodies, which appear yellow when they overlap. This is problematic, however, if the intensities of the probes are not similar to one another, in this case one color may dominate the other and the intermediate yellow cannot be seen with the naked eye[94]. In order to circumvent subjective visualization issues, the first colocalization analysis tool employed in this project was the graphic projection of the z-axis profile in specific cellular regions of the two fluorophores labeling the proteins of interest (see sections 2.14.3, 3.4.3, and 3.9.3).

It is also possible to quantify colocalization. The main method is the calculation of the Pearson's correlation coefficient, which indicates the degree of colocalized signal in both channels[95,106]. This calculation, however, measures pixel intensity correlation, and does not account for possible differences in fluorescence intensity, which would result in low correlation coefficients for two fluorophores that are spatially overlapping but have different intensities. To correct for any differences in labeling intensities between GLuc and the potential colocalized structure of interest, the Overlap correlation coefficient was utilized to assess colocalization in this project. This calculation removes the mean intensity value of both analyzed channels in order to determine actual overlap of the two channels in a particular voxel[93,106]. In this project, the colocalization analysis is depicted using a representative confocal image with a region of interest highlighted. From this region, the z-axis fluorophore profile of the proteins of interest are plotted. This data is supplemented with the overlap coefficient for each region of interest.

Proof of concept – colocalization analysis of neurosecretory vesicles by confocal microscopy

As elaborated in the Introduction (section 1.2) and mentioned above, neurosecretory vesicle-associated proteins travel through the Golgi stack, where they are sorted into immature vesicles which bud off of the *trans*-Golgi network[15]. All of the clones in this project described in detail above, except for the negative control no tag GLuc clone, express GLuc fused to a sorting signal or protein associated with either large dense core vesicles or synaptic vesicles. Therefore, in order to verify that the fusion proteins are not only being properly expressed, but are properly expressed in the correct location, the colocalization analysis was carried out by comparing the localization of GLuc with the localization of either the Golgi apparatus (in iPSCs) or the appropriate neurosecretory vesicle (in SIMA clones and differentiated MNs). Proteins associated with both LDCVs and SVs, for example CgA and VAChT, have been shown by overlap correlation coefficient analysis of confocal images to be colocalized with the Golgi

apparatus[17]. Furthermore, the scrapie responsive gene one protein was found not only to be highly colocalized with the Golgi apparatus, but it was also highly associated with SgII, indicating that this protein was moving through the neurosecretory pathway into LDCVs[157]. Additionally, the glucose transporter GLUT4 could be localized by immunofluorescence microscopy to the perinuclear region and distal processes of differentiated PC12 cells and was confirmed to colocalize with the SV marker protein synaptophysin[158]. Finally, the calcium-sensor protein synaptotagmin 7 was shown in insulin-secreting cells to not colocalize with LDCVs, but rather colocalizes with Rab7 on endosomes[159]. Given these examples, it is appropriate to detect whether GLuc is colocalized with the Golgi, LDCVs, SVs, or none of the structures, using confocal microscopy and colocalization coefficient analyses.

Colocalization analysis – clones associated with LDCV signal peptides

In the SIMA hPOMC-GLuc clone, GLuc had a high degree of colocalization with the LDCV protein CgA[20] and in the IMR90-4 hPOMC-GLuc clones 4 and 6, GLuc had a high degree of colocalization with the Golgi apparatus. Therefore, GLuc is sorted correctly in the IMR90-4 clones, with the expectation that, once differentiated, they can successfully transport the newly packaged GLuc from the Golgi apparatus to the presynaptic terminal in LDCVs[16,17].

Both IMR90-4 CgA-GLuc and IMR90-4 SgII-GLuc clones are made of constructs that, like hPOMC, are designed to deliver GLuc to LDCVs. In the IMR90-4 CgA-GLuc clone 8, the protein labeling associated with GLuc was visible adjacent to the nucleus in a region close to the Golgi apparatus, but it did not replicate the highly organized elongated accumulation identified in the IMR90-4 hPOMC-GLuc clones. Instead, within this area, GLuc aggregates appeared very coarse and irregular, in some areas the staining appears faint and diffuse and at others globular and amassed (Figure 80 and Figure 81). This may point to an abnormal degradation of GLuc when it is associated with the CgA sorting signal. Protein aggregates can also be identified by IF in cells in which proteasomal and autophagic protein degradation is inhibited[160]. Furthermore, misfolded proteins have been shown to accumulate next to but not in the Golgi apparatus. The misfolding could be caused by errors in transcription, mRNA processing, or translation[161].

Both CgA and SgII sorting sequences are 2.5 - 3.6 times larger, respectively, than the GLuc protein (section 6.3). In both cases, the problematic GLuc sorting and expression could be due to interference of the very large targeting sequences. Presecretory signal peptides have been shown to vary in length and amino acid composition, but have been shown to share key

176

features, such as a positively charged residue preceding a hydrophobic core approximately 12 AAs long[162]. All three LDCV targeting sequences contain these features (see section 6.3.2), however the hPOMC signal is only 26 amino acids long, whereas CgA and SgII range from 450 to 620 AAs, with their presecretory signal peptide appearing in the first 25 AAs of the sequence. Furthermore, signal peptidase recognition sites (preceded by alanine-X-alanine) occur often within several AAs of the signal peptide[162]. Therefore, if the cleavage of the CgA and SgII signal peptides takes place at the beginning of these long sequences, the sequence still attached to GLuc may be interfering with protein folding. There was no further analysis on the colocalization of GLuc in the IMR90-4 CgA clones. The IMR90-4 SgII-GLuc clone 11 only expressed very low levels of luciferase, as seen in both the initial luciferase activity test and in IF, which was did not consistently and clearly detect GLuc (Figure 82). For this reason, the IMR90-4 SgII-GLuc clones were not used in any further experiments.

Colocalization analysis – clones associated with the SV protein VAMP2

In the immunofluorescence analysis of the SIMA VAMP2-GLuc clone 1 only faint staining associated with GLuc could be seen throughout the cell. There was little to no visible sorting through the Golgi or colocalization with synaptophysin, a protein typically associated with SVs (Figure 83). On the other hand, the IMR90-4 VAMP2-GLuc clone 11 did show colocalization of GLuc with the Golgi, albeit with a lower correlation than that of hPOMC-GLuc with either CgA (in SIMA cells) or Golgi (in IMR90-4 cells), but also with a significantly higher correlation than no tag GLuc (Figure 84). The decrease in colocalization could be due to complications originating from the construct itself: the Gaussia luciferase protein is 19.9 kDa[86], while the VAMP2 protein is only 13 kDa, with only a very short C-terminus amino acid tail entering the SV lumen[24,87]. The GLuc, attached to this short amino acid tail destined for the lumen, is therefore even larger than the VAMP2 protein. This likely effects the behavior and folding of VAMP2, as well as the transmembrane insertion of GLuc into the SV lumen.

Proof of concept – reaching the synaptic vesicle utilizing VAMP2

Other groups have successfully constructed fusion proteins with VAMP2, on both the N and C termini, and have confirmed their final arrival at the vesicle membrane[55,87]. Therefore, the fusion protein described here should theoretically sort to the synaptic vesicles. Examples of VAMP2-associated fusion proteins include one in which green fluorescent protein (GFP) was fused to the VAMP2 N-terminus and expressed in SIMA cells. The correct sorting of GFP was proven by cellular imaging, whereby under standard conditions the GFP was seen to be

associated with vesicular structures, and after exposure to BoNT/B, the GFP was cleaved from the VAMP2 protein and could be seen throughout the cytosol[55]. The fusion of GFP to the N terminus of VAMP2 likely increases the chances of successful sorting to the neurosecretory vesicles, since the majority of VAMP2 is localized on the cytosolic side of the vesicle and no translocation of the reporter protein into the vesicular lumen must take place. However another group successfully fused YFP (yellow fluorescent protein) and GFP, respectively, to the N and C termini of VAMP2 in order to investigate the mechanism of protein sorting to SVs. The cellular localization of these fusion proteins was also analyzed by fluorescence microscopy and the two fusion proteins were found to have similar vesicular expression patterns in hippocampal neurons[87].

In both examples, the verification of the fusion protein sorting onto the neurosecretory vesicle was only analyzed by fluorescence microscopy detecting the specific reporter protein. No quantifiable colocalization analyses were completed, rather the enrichment of fluorescence was observed around vesicle-like puncta in the axons or surrounding the nucleus in a Golgi-like manner. Therefore the degree of colocalization as analyzed in this report is unique to the analysis of VAMP2 fusion protein constructs and the degree of successful sorting cannot be compared directly to these previously reported constructs. Likely, however, in the case of the IMR90-4 VAMP2-GLuc clone 11, a fraction of the fusion protein is successfully reaching the SV, while the remaining fraction is either found in the LDCV or missorted. This point must be kept in mind while testing the release of luciferase in the MoN-Light BoNT assay, because the measurement of luciferase upon depolarization must be due to the fusion of the SV to the presynaptic membrane and not due to the permanent availability of a possibly missorted fusion protein.

4.4 iPSC derived motor neurons

4.4.1 Establishment of motor neuron differentiation

The tests to optimize the differentiation protocol were performed in a parallel PhD thesis by a cooperating group at the University of Veterinary Medicine in Hannover[99,163]. In short, three peer-reviewed protocols developed by Du *et al*, Maury *et al*, and Kroehne *et al* were implemented and analyzed in non-transfected IMR90-4 cells and hPOMC-GLuc clones. All three protocols produced mixed motor neuron populations. The Du *et al* protocol could achieve differentiated cell populations containing approximately 50 % motor neurons, but these results

178

varied considerably (SD $\pm$ 25 %). Both the Maury *et al* and Kroehne *et al* protocols more consistently resulted in differentiated cell populations with only approximately 15 % motor neurons. All protocols were shown to produce cells expressing all the receptors and substrates necessary for BoNT sensitivity[99].

Before the final motor neuron population analysis could be completed it was communicated that the easiest protocol to carry out, and therefore the best candidate for the MoN-Light BoNT assay, was that described by Maury *et al*. Therefore this was the first protocol used to test the potential response of the IMR90-4 hPOMC-GLuc clone in the luciferase release assay. Only later was it demonstrated that the protocol developed by Du *et al* resulted in the highest percentage of MNs in the total pool of cultured cells, indicating that this protocol might actually be more appropriate for the preparation of cells for the MoN-Light BoNT assay. Therefore the Du *et al* protocol was later applied in the IMR90-4 hPOMC-GLuc, IMR90-4 no tag GLuc, and IMR90-4 VAMP2-GLuc clones. Unfortunately, the IMR90-4 hPOMC-GLuc clones all died approximately 2 – 3 weeks into the differentiation in two independent experiments. These clones were differentiated parallel to the IMR90-4 VAMP2-GLuc and IMR90-4 no tag GLuc clones, which were successfully differentiated into a population containing motor neurons (see section 3.8). Therefore any issues with a problematic batch of medium can be excluded.

The Maury *et al* differentiation protocol involves only two major handling steps: the cells first grow into neurospheres in suspension, and then they are directly plated on day 9 onto Matrigel-coated plates until day 30[9,163]. On the other hand, the Du *et al* protocol involves splitting steps at day 6 and again at day 12, at which point the cells grow into neurospheres in suspension. Finally at day 18 the cells are plated onto Matrigel-coated plates for final motor neuron maturation[8,163]. Since the IMR90-4 hPOMC-GLuc clone 6 could be successfully differentiated according to Maury *et al*, but not according to Du *et al*, the cells might be more sensitive to the multiple and stressful handling steps carried out in the latter protocol. Clonal variability can arise and affect the differentiation process[164], so although the CRISPR-modified IMR90-4 no tag GLuc and IMR90-4 VAMP2-GLuc clones could be successfully differentiated according to Du *et al*, this does not preclude the successful differentiation of the IMR90-4 hPOMC-GLuc clone. No problems with cell survival arose during the Maury *et al* differentiation of any of the clones.

4.4.2　iPSCs and the motor neuron differentiation

Confirmation of IMR90-4 pluripotency

Pluripotency can be summarized as the stem cells' expression of pluripotency markers, their ability to self-renew, and their ability to differentiate into all three embryonic germ layers[165]. Gene expression of the four genes, *SOX2*, *OCT4*, *NANOG* and *LIN28,* used to reprogram the IMR90-4 cells from fetal lung fibroblasts into iPSCs[166], was measured to confirm the clones' undifferentiated state. The non-transfected IMR90-4 cells and the clones relevant for the MoN-Light BoNT assay (IMR90-4 hPOMC-GLuc clone 6 and 4, and IMR90-4 VAMP2-GLuc clone 11) all expressed these genes, while one MN sample analyzed alongside the iPSCs as a control no longer expressed any of the four genes (Figure 93). It is important to verify the pluripotent status of the non-differentiated cells to exclude possible spontaneous differentiation before the implementation of the differentiation protocol. For example, Nanog protein expression is essential for preserving pluripotency and its lowered expression can induce unwanted differentiation[127]. During the investigation of embryoid body based spontaneous differentiation of over 60 iPS cell lines, it was determined that lower expression of pluripotency-associated genes, as compared to embryonic stem cells, resulted in differentiated clones with more variable phenotypes. It was concluded that iPSCs can be pre-selected for more consistent differentiation characteristics based upon this transcriptional screen[164]. Since the IMR90-4 cell line was purchased from a cell repository and had previously been evaluated for their ability to form the endoderm, mesoderm, and ectoderm, it was considered sufficient to confirm the pluripotent status by the observation of the cell line's ability to continuously proliferate and to express the appropriate pluripotent markers.

Confirmation of motor neuron differentiation

The end point of the differentiation can be assessed by the expression of genes associated with the presence of enriched motor neurons, *ISLET1*, *HB9*, and *CHAT*[8,9]. *ISLET1* and *HB9* are essential transcription factors in motor neuron generation, and *CHAT* is responsible for the synthesis of acetylcholine[13,167]. Islet1 has been shown to direct cholinergic neuron generation by forming multi-protein complexes which bind to and promote the expression of genes critical for acetylcholine synthesis and packaging, such as *CHAT* and *VAChT*. When the expression of *ISLET1* is triggered, cholinergic pathway genes such as *CHAT* were shown to be triggered, while the expression of MN gene *HB9* was unaffected[168]. Global transcriptome profiling of *in vitro* motor neuron differentiation has shown by RT-PCR that the expression of

180

each of these tissue development markers after 18 to 28 days of differentiation are similar to one another[167]. In the original descriptions of the differentiation methods tested here, the presence of the markers after 30 days of differentiation was only confirmed by immunofluorescence and percent of marker positive cells, but the expression was not quantified nor compared between genes[8,9], therefore it is not possible to compare expression levels of MNs differentiated for this project with those from the original protocols.

The differentiation of IMR90-4 hPOMC-GLuc clone 6 according to Maury *et al* assessed at day 30 resulted in cell populations expressing *ISLET1*, *HB9*, and *CHAT*. The IMR90-4 no tag GLuc clone 4 differentiation according to Du *et al* resulted in cells highly expressing *ISLET1*, but only marginally expressing *HB9* and *CHAT*. IMR90-4 VAMP2-GLuc clone 11 was successfully differentiated according to both Du *et al* and Maury *et al* protocols, the derived populations both highly expressed all three motor neuron markers (Figure 94). These results confirm that at least a portion of the cells in each of the executed differentiations were successfully transformed into motor neurons, with potentially higher efficiency in IMR90-4 hPOMC-GLuc and IMR90-4 VAMP2-GLuc clones than in the IMR90-4 no tag GLuc clone. Gene expression analyses were only carried out once for the IMR90-4 no tag GLuc clone, so the lack of replicate data for this clone may have introduced an error into the results, making this data point more variable and difficult to interpret.

Confirmation of GLuc expression in motor neurons

Importantly, all the clones, whether differentiated or not, expressed GLuc. However, the differentiation process did reduce, but did not abolish, total GLuc expression in each of the clones tested (Figure 94). Several cells could be identified by DAPI that were not expressing GLuc (see section 3.9.1). Some of these DAPI labeled cells, as mentioned in the results, are apoptotic. However it does appear that some live cells do not express GLuc. The clones have been isolated and expanded to be monoclonal populations, which can be seen in the IF analysis of GLuc expression in the iPSCs (see sections 3.2.3.2, 3.4.3, and 6.6.3.2). However, there is a possibility that a very small number of puromycin-resistant cells that do not express GLuc are hidden in this population. On the other hand, given the decrease of GLuc expression post-differentiation, there is also the possibility that a sub-section of the differentiated cells stop expressing GLuc. It is therefore important to verify that those cells still expressing GLuc are also part of the differentiated motor neuron population.

The combination of the expression of GLuc and the motor neuron markers is essential for this project. Gene expression qPCR analysis only provides general information for the entire population of cells analyzed, but cannot determine if exactly those cells which express GLuc are also those which express the MN markers. Therefore, immunofluorescence is once again very demonstrative in the characterization of the MNs. The expression of GLuc could be confirmed in the differentiation according to Maury *et al* for both IMR90-4 hPOMC-GLuc and IMR90-4 VAMP2-GLuc, as well as the differentiation according to Du *et al* for IMR90-4 VAMP2-GLuc (Figure 95). Furthermore the expression of GLuc in Islet1 positive motor neurons could be confirmed in all three differentiation scenarios (Figure 96 through Figure 98). This evidence is essential to address the question of whether GLuc is actually expressed in the motor neurons, and if not, if that might prevent the IMR90-4 hPOMC-GLuc MNs from reliably releasing luciferase. The confirmation of GLuc expression in motor neurons excludes this factor from the potential problematic aspects of the MoN-Light BoNT assay.

Confirmation of GLuc localization in LDCVs and SVs

While it is essential that GLuc can be found in motor neurons, it is also important to verify that the protein is correctly sorted into vesicles from the Golgi apparatus post-differentiation. The colocalization of GLuc in the IMR90-4 hPOMC-GLuc differentiated cells was evaluated against the LDCV associated protein SgII[20,137], and in the IMR90-4 VAMP2-GLuc differentiated cells against the SV associated protein Syp[153,169] (Figure 99 through Figure 101). The overlap coefficient for each clone was not significantly different from each clone's corresponding degree of correlation of GLuc with the Golgi apparatus (Figure 102). It appears that the same degree of protein that was correctly sorted in the pluripotent cells through the Golgi is also successfully sorted in the differentiated cells into the neurosecretory vesicles. This verification also supports the assumption that both clone constructions should theoretically be capable of releasing luciferase upon cellular depolarization.

4.5 MoN-Light BoNT Assay

4.5.1 Luciferase activity in supernatant upon depolarization

Verification of luciferase release from CRISPR-modified SIMA clones

The luciferase release assay is described in full in section 1.6.3. All SIMA clones contained off-target donor DNA insertions and were not the perfect controls for the luciferase

release assay. Despite this, a selection of SIMA hPOMC-GLuc clones were tested for their capacity to release luciferase upon cellular depolarization. The standard preliminary release experiment with control and depolarization buffers both in the absence and presence of EGTA was carried out with the SIMA IMR90-4 hPOMC-GLuc clones 1, 2, 4, and 5. Clone 1 was shown to have at least ten copies of GLuc inserted in its genome, clone 2 was shown to have approximately four copies of GLuc, and clones 4 and 5 each had one on-target GLuc insert and one off-target GLuc insert (Figure 67). Exposure to the depolarization buffer increased luciferase activity in the supernatant by approximately 3-4 fold and the activity increase could be completely suppressed in combination with EGTA (Figure 89). These results are in accordance with those of the SIMA Random-Insertion_hPOMC1-26GLuc prototype[2], which was also shown to have approximately 4 copies of GLuc inserted randomly into its genome (Figure 67). In this case, the difference in genetic editing method and copy number of donor DNA in the genome did not have a great effect on the outcome of the release of luciferase into the buffer. This confirms that hPOMC-GLuc clones derived by the CRISPR/Cas9 genetic editing technique are capable of releasing the GLuc that has been sorted into the LDCV upon cellular depolarization.

Luciferase release from IMR90-4 hPOMC-GLuc derived motor neurons

The IMR90-4 hPOMC-GLuc clone 6 derived MNs differentiated according to Maury *et al* could not be provoked to release luciferase into the cell culture supernatant. Furthermore, upon exposure to control and depolarization buffers including EGTA, which should suppress luciferase release, even greater amounts of luciferase activity could sometimes be measured in the supernatant, resulting in extremely high variation between tests (Figure 103). As a negative control, the IMR90-4 no tag GLuc clones were also differentiated into motor neurons and tested for luciferase release. Exposure to control and depolarization buffers, in addition to EGTA resulted in similar extremely variable luciferase activity in the supernatant (Figure 103). The high variability between tests, especially in the presence of EGTA is as of yet unexplained. The sequestration of Ca^{2+} by EGTA should not have stimulated luciferase release. This confirms that the SIMA luciferase release protocol cannot be transferred to motor neurons because it yields non-reproducible random results.

4.5.2 Optimization of MoN-Light BoNT Assay

4.5.2.1 Modifications to IMR90-4 hPOMC-GLuc derived motor neuron release protocol

In order to optimize potential hPOMC-GLuc luciferase release patterns, an alternative pre-incubation medium and the seeding of a range of MN cell densities was tested (Figure 104). The alternative pre-incubation medium was tested in order to exclude the possibility that the cell culture medium was not already stimulating and exhausting the cells and therefore eliminating luciferase release during exposure to depolarization buffer. The pre-incubation buffer N2B27, as described in the Materials and Methods (section 2.15), consists of a 50/50 mixture of DMEM/F12 medium (ThermoFisher #21331020[170]) and Neurobasal medium (ThermoFisher #21103049[171]) with N2 and B27 supplements, which is also the basis buffer in which the motor neurons were cultivated during their differentiation. The DMEM/F12 medium, however, contains glutamic acid, an excitatory neurotransmitter in the brain[172]. It was hypothesized that the Neurobasal medium, which does not contain glutamic acid, might be a more appropriate pre-incubation medium. However, after pre-incubation with Neurobasal medium, there was still no specific luciferase release after exposure to the depolarization buffer and the buffers including EGTA once again produced an extremely variable increase of luciferase activity in the supernatant.

The type of medium and its components in which cultured neuronal cells are incubated has been proven to be important for cellular excitotoxicity[173,174]. Confirming the initial hypothesis that DMEM/F12 medium was exhausting the cells' ability for synaptic release, it has been shown that DMEM/F12 medium consistently depolarizes the resting potential of neurons and can saturate and silence the firing of the cells[175]. Excitatory amino acids were specifically removed from the Neurobasal medium, which was shown to not depolarize the resting membrane potential. However, this medium was shown to reduce voltage-dependent sodium currents and rapidly inactivating potassium currents, and therefore also debilitated both evoked and spontaneous action potentials[175]. The same group reporting this data developed a novel BrainPhys medium, which enhances neuronal synaptic function and may be a good solution to improve the electrophysiological conditions for the MoN-Light BoNT assay.

A series of decreasing densities of MNs were plated in preparation for the luciferase release test in order to investigate the influence of confluency on the degree of luciferase released into the cell culture supernatant. It has been shown that neurite outgrowth and cell

survival can be influenced by the cell density of cultured neurons[176–178]. However, the series of luciferase release tests with reduced MN confluencies resulted in the smallest differences in luciferase activity between any of the buffers, probably due to the fact that there were too few cells to elicit any type of cellular response.

In summary, the IMR90-4 hPOMC-GLuc derived MNs, in combination with the existing standard depolarization protocol, are not yet optimized for use in the MoN-Light BoNT assay. The obstacle with using these cells does not originate from the construct or the genetic editing technique, proven by the proper functionality of SIMA clones prepared with the same techniques, and the extensive IF colocalization characterization of the IMR90-4 derived cells. Rather, the inability of the MNs to respond to external stimuli as expected is probably due to the fact that the cellular depolarization protocol was optimized for SIMA clones and is not appropriate for MNs.

4.5.2.2 Improved applicability of GLuc vesicle packaging: synaptic vesicles

Rationale for synaptic vesicle construct

The functionality of the MoN-Light BoNT assay is completely dependent on the packaging of GLuc in neurosecretory vesicles. The co-release of neurotransmitters and GLuc from neurosecretory vesicles at the presynaptic terminal into the buffer surrounding the cell is the key factor in discerning normally functioning cells (under no or very low exposure to BoNT) from paralyzed cells (under exposure to BoNT). As discussed in the Introduction (1.2.1), both LDCVs and SVs are found in the presynaptic terminal of motor neurons[153]. Although the membrane fusion of both vesicle types is regulated by the proteins forming the SNARE complex, the trigger for fusion comes from different signals, with the LDCVs associated with slow release upon prolonged stimulation, while SVs can fuse and release their contents in response to a single action potential[23,179]. The standard and most reliable method to induce action potentials is through depolarizing current injections by intracellular microinjections[23,179]. This method is, however, very specialized and not conducive to the goal of the MoN-Light BoNT project to develop an easily applicable toxicity assay. The exposure of the cells to the depolarization buffer described for the luciferase release experiment causes a quick influx of Ca^{2+} into the cell more closely mimicking a single action potential, and therefore would more likely trigger the release of neurotransmitter-containing SVs rather than LDCVs[115,179]. The VAMP2-GLuc construct was designed to potentially address the issue of differential vesicle release. Since VAMP isoforms are found on both vesicle types[87,180–182], the chance of luciferase

release provoked by the depolarization buffer is greater and any differences in the evocation of vesicular fusion events with the presynaptic membrane should be irrelevant. Furthermore, the association of VAMP2-GLuc with SVs, in which the neurotransmitters are stored, increases the biological relevance of the assay, precisely tracking neurotransmitter release and therefore neuronal signal propagation.

As previously discussed, the construction of the VAMP2-GLuc clones fuses GLuc to the vesicular transmembrane protein VAMP2 with a TEV recognition sequence linker. Due to its small size and natural abundance, the SV has been very well characterized. These vesicles are only known to transport neurotransmitters and the currently known proteins found with the vesicle are all membrane-associated. Therefore it was impossible to take advantage of a signal peptide of a protein that would be packaged into the SV, since they are all bound to the vesicle[26,83,84,153,169]. The only known protein with the C-terminus ending in the SV lumen is VAMP, and was therefore the clear choice with which to transport GLuc into the lumen. The fusion of GLuc to the transmembrane protein does, however, create one complication: if GLuc cannot move freely in the solution surrounding the cell, it is impossible to collect and measure the luciferase activity in the supernatant. Therefore the TEV recognition sequence linker[85] was included to provide the opportunity to free GLuc from VAMP during fusion of the SV to the presynaptic membrane upon exposure to the TEV protease (see 1.6.1.3 for TEV protease cleavage schematic).

4.5.2.3 Luciferase release from IMR90-4 VAMP2-GLuc derived motor neurons

Luciferase release protocol for IMR90-4 VAMP2-GLuc derived motor neurons

The luciferase release protocol was drastically modified for the IMR90-4 VAMP2-GLuc clone, due to the fact that the VAMP2-GLuc fusion protein must be cleaved by the TEV protease in order to collect free luciferase in the supernatant. The optimal temperature for TEV protease activity is between 4 - 30 °C[113], so the first release experiment was tested at 30 °C for 20 min, in order to allow the 5 U of TEV protease to have more time to cleave the VAMP2-GLuc fusion protein. This experimental setup did not result in any identifiable increase of luciferase in the cell culture supernatant (Figure 105). Therefore, a wider range of TEV protease was added to the respective buffers to determine the best enzyme concentration to cleave VAMP2-GLuc. Furthermore, the cells were incubated with the protease for 2 h at 24 °C. The cells were incubated for 2 h in order to allow cleaved luciferase to accumulate in the

supernatant. The cells were incubated at room temperature because it has been shown that the mean rate of endocytosis of synaptic vesicles at physiological temperatures is significantly faster than the mean rate of endocytosis at room temperature in rat calyx of Held nerve terminals[114]. Therefore, it was hypothesized that the availability of the fusion protein to the TEV protease would increase and consequently have a greater chance of successful cleavage if vesicle endocytosis were delayed. Indeed, these modifications to the release protocol resulted in significantly higher luciferase activity in the cell supernatant with 10 U TEV protease than without (Figure 106). After optimizing the control/depolarization buffer incubation length and temperature, and finding the appropriate TEV protease concentration, a series of experiments were carried out in IMR90-4 VAMP2-GLuc cells differentiated according to both Maury *et al* and Du *et al*.

Luciferase release from IMR90-4 VAMP2-GLuc derived motor neurons after exposure to BoNT/A

The IMR90-4 VAMP2-GLuc derived MN population differentiated according to Du *et al* was the only population submitted to the MoN-Light BoNT test. Preliminary experiments had shown that exposure of the differentiated cells to TEV protease successfully resulted in an increase of luciferase in the cell culture supernatant. Therefore, this experimental conformation was tested with BoNT/A. Half of the cells were cultivated normally, while the other half were incubated for 48 hours with BoNT/A. The release test was carried out upon exposure to medium, control buffer, and depolarization buffer, each with and without the TEV protease. The cells exposed to BoNT/A and then incubated with TEV protease in Na^+-HBS buffer still produced an increase in luciferase activity in the supernatant in comparison to incubation with Na^+-HBS buffer alone. This increase was, however, significantly lower than the increase of luciferase activity after incubation with Na^+-HBS buffer and TEV protease without exposure to BoNT/A (Figure 107). This indicates that a portion of the luciferase activity detected after incubation of Na^+-HBS buffer and TEV protease is likely due to the fusion of synaptic vesicles to the presynaptic membrane and the cleavage of GLuc into the buffer surrounding the cells. This fusion is then blocked after exposure to BoNT/A and the subsequent cleavage of SNAP-25 and disruption of the SNARE complex[24,26,181].

Further optimization of luciferase release protocol for IMR90-4 VAMP2-GLuc derived motor neurons

Differentiated MNs were exposed to N2B27 medium, Na^+-HBS buffer, and K^+-HBS buffer including 10 U TEV protease and in the absence and presence of 10 mM EGTA and

1 mM carbachol. Carbachol is a synthetic derivative of choline and can stimulate muscarinic and nicotinic receptors to release acetylcholine at certain neuromuscular junctions[183]. Furthermore carbachol was shown to strongly increase Ca^{2+} accumulation[184], possibly by activating an Na^+-Ca^{2+} exchanger[185]. The compound has also been shown in crayfish to cause rhythmic bursts of motor neuron activity[186]. Na^+-HBS buffer with TEV protease and carbachol could elicit a significantly higher luciferase release than Na^+-HBS buffer with TEV protease alone from MNs differentiated according to both protocols. This significant increase in luciferase activity in the supernatant could be significantly reduced, though not completely eliminated, in the presence of EGTA (Figure 109 and Figure 110). The carbachol-elicited release from the IMR90-4 VAMP2-GLuc derived MNs was compared to the IMR90-4 no tag GLuc derived MNs. The specificity of the luciferase release from the VAMP2-GLuc clone after stimulation with carbachol could be verified by Sidak's multiple comparison test (Figure 111). The combination of the reduction of carbachol-elicited luciferase release after exposure to EGTA and the specificity of the release as compared to the no tag GLuc clone build a strong argument that the exposure to carbachol successfully stimulates vesicle fusion to the presynaptic membrane of IMR90-4 VAMP2-GLuc derived MNs. Upon vesicle fusion, the TEV protease can access and cleave its recognition site, thus freeing GLuc from VAMP2, and releasing it into the cellular supernatant. The potential applicability of the addition of carbachol as a stimulant of vesicular fusion to the pre-synaptic membrane is very high and must still be implemented in the VAMP2-GLuc MoN-Light BoNT assay in order to fully validate the hypothesis that the compound does in fact instigate exocytosis.

4.6 Conclusions

While the actual end-goal of this project, the development of an accurate and sensitive *in vivo* simulation assay to measure the potency of BoNT was not achieved, an important portion of the project was successfully accomplished. Two strong candidates to be used in the cell-culture assay were generated and thoroughly characterized in both iPSC and MN states. The clones were accurately modified using the CRISPR/Cas9 genetic engineering method, and were proven to contain homozygote copies of the Gaussia luciferase reporter sequence in the AAVS1 safe harbor locus. The double-control quantitative copy number PCR method to detect possible off-target donor DNA insertions was developed and utilized to verify that the assay candidates did not contain any off-target modifications. Finally, the localization of luciferase reporter protein in the iPSCs and MNs was verified in the Golgi apparatus and the

188

neurosecretory vesicles, to ensure that the clones should be functionally capable of releasing their vesicular contents into the cell culture supernatant during the assay. Furthermore, the use of carbachol as a stimulant of vesicular exocytosis lends even greater potential to the future functionality of this assay, but must also be authenticated and optimized. The provoked co-release of neurotransmitters and luciferase has not yet been perfected, but the system is at the stage to undergo testing of a wide-range of factors to establish a functional protocol.

5 References

1. Bitz, S. The botulinum neurotoxin LD50 test - problems and solutions. *ALTEX* **27**, 114–116 (2010).

2. Pathe-Neuschäfer-Rube, A., Neuschäfer-Rube, F., Genz, L. & Püschel, G. Botulinum neurotoxin dose-dependently inhibits release of neurosecretory vesicle-targeted luciferase from neuronal cells. *ALTEX* (2015) doi:10.14573/altex.1503061.

3. The Initial Formation of the Nervous System: Gastrulation and Neurulation. in *Neuroscience* (eds. Purves, D., Augustine, G., Fitzpatrick, D. & et al) (Sinauer Associates, 2001).

4. Song, M.-R. & Pfaff, S. L. Motor Neuron Specification in Vertebrates. in *Encyclopedia of Neuroscience* 1015–1022 (Elsevier, 2009). doi:10.1016/B978-008045046-9.01035-4.

5. The Molecular Basis of Neural Induction. in *Neuroscience* (eds. Purves, D., Augustine, G., Fitzpatrick, D. & et al) (Sinauer Associates, 2001).

6. Stifani, N. Motor neurons and the generation of spinal motor neuron diversity. *Front. Cell. Neurosci.* **8**, (2014).

7. Perrimon, N., Pitsouli, C. & Shilo, B.-Z. Signaling Mechanisms Controlling Cell Fate and Embryonic Patterning. *Cold Spring Harb. Perspect. Biol.* **4**, a005975–a005975 (2012).

8. Du, Z.-W. *et al.* Generation and expansion of highly pure motor neuron progenitors from human pluripotent stem cells. *Nat. Commun.* **6**, 6626 (2015).

9. Maury, Y. *et al.* Combinatorial analysis of developmental cues efficiently converts human pluripotent stem cells into multiple neuronal subtypes. *Nat. Biotechnol.* **33**, 89–96 (2015).

10. Little, D., Ketteler, R., Gissen, P. & Devine, M. J. Using stem cell–derived neurons in drug screening for neurological diseases. *Neurobiol. Aging* **78**, 130–141 (2019).

11. Brady, S. T., Colman, D. R. & Brophy, P. J. Subcellular Organization of the Nervous System. in *Fundamental Neuroscience* 61–92 (Elsevier, 2013). doi:10.1016/B978-0-12-385870-2.00004-4.

12. Felten, D. L., O'Banion, M. K. & Maida, M. S. Neurons and Their Properties. in *Netter's Atlas of Neuroscience* 1–42 (Elsevier, 2016). doi:10.1016/B978-0-323-26511-9.00001-1.

13. Sha, D., Jin, H., Kopke, R. D. & Wu, J.-Y. Choline acetyltransferase: regulation and coupling with protein kinase and vesicular acetylcholine transporter on synaptic vesicles. *Neurochem. Res.* **29**, 199–207 (2004).

14. Hayworth, C. R. Chapter 5 - The Formation and Maturation of Neuromuscular Junctions. 23.

15. *Neurosecretion: Secretory Mechanisms.* vol. 8 (Springer International Publishing, 2020).

16. Deutch, A. Y. Neurotransmitters. in *Fundamental Neuroscience* 117–138 (Elsevier, 2013). doi:10.1016/B978-0-12-385870-2.00006-8.

17. Park, J. J., Gondre-Lewis, M. C., Eiden, L. E. & Loh, Y. P. A distinct trans-Golgi network subcompartment for sorting of synaptic and granule proteins in neurons and neuroendocrine cells. *J. Cell Sci.* **124**, 735–744 (2011).

18. Stenoien, D. & Brady, S. T. Fast Axonal Transport. in *Basic Neurochemistry: Molecular, Cellular and Medical Aspects* (eds. Siegel, G., Agranoff, B. & Albers, R.) (Lippincott-Raven, 1999).

19. Takamori, S. *et al.* Molecular Anatomy of a Trafficking Organelle. *Cell* **127**, 831–846 (2006).

20. Bartolomucci, A. *et al.* The Extended Granin Family: Structure, Function, and Biomedical Implications. *Endocr. Rev.* **32**, 755–797 (2011).

21. Chevrier, D. *et al.* Expression of porcine pro-opiomelanocortin in mouse neuroblastoma (Neuro2A) cells: Targeting of the foreign neuropeptide to dense-core vesicles. *Mol. Cell. Endocrinol.* **79**, 109–118 (1991).

22. Sieburth, D., Madison, J. M. & Kaplan, J. M. PKC-1 regulates secretion of neuropeptides. *Nat. Neurosci.* **10**, 49–57 (2007).

23. Bruns, D. & Jahn, R. Real-time measurement of transmitter release from single synaptic vesicles. *Nature* **377**, 62–65 (1995).

24. Montecucco, C., Schiavo, G. & Pantano, S. SNARE complexes and neuroexocytosis: how many, how close? *Trends Biochem. Sci.* **30**, 367–372 (2005).

25. Jahn, R. Principles of exocytosis and membrane fusion. *Ann. N. Y. Acad. Sci.* **1014**, 170–178 (2004).

26. Rizo, J. & Südhof, T. C. Snares and munc18 in synaptic vesicle fusion. *Nat. Rev. Neurosci.* **3**, 641–653 (2002).

27. Südhof, T. C. Neurotransmitter Release: The Last Millisecond in the Life of a Synaptic Vesicle. *Neuron* **80**, 675–690 (2013).

28. Kumar, R., Dhaliwal, H., Kukreja, R. & Singh, B. The Botulinum Toxin as a Therapeutic Agent: Molecular Structure and Mechanism of Action in Motor and Sensory Systems. *Semin. Neurol.* **36**, 010–019 (2016).

29. Pantano, S. & Montecucco, C. The blockade of the neurotransmitter release apparatus by botulinum neurotoxins. *Cell. Mol. Life Sci.* **71**, 793–811 (2014).

30. Pirazzini, M., Rossetto, O., Eleopra, R. & Montecucco, C. Botulinum Neurotoxins: Biology, Pharmacology, and Toxicology. *Pharmacol. Rev.* **69**, 200–235 (2017).

31. Schantz, E. J. & Johnson, E. A. Properties and use of botulinum toxin and other microbial neurotoxins in medicine. *Microbiol. Rev.* **56**, 80–99 (1992).

32. Pickett, A. Botulinum Toxin as a Clinical Product: Manufacture and Pharmacology. in *Clinical Applications of Botulinum Neurotoxin* (ed. Foster, K. A.) 7–49 (Springer New York, 2014). doi:10.1007/978-1-4939-0261-3_2.

33. Wortzman, M. S. & Pickett, A. The science and manufacturing behind botulinum neurotoxin type A-ABO in clinical use. *Aesthet. Surg. J.* **29**, S34-42 (2009).

34. Montecucco, C. & Schiavo, G. Structure and function of tetanus and botulinum neurotoxins. *Q. Rev. Biophys.* **28**, 423–472 (1995).

35. Montal, M. Botulinum Neurotoxin: A Marvel of Protein Design. *Annu. Rev. Biochem.* **79**, 591–617 (2010).

36. Turton, K., Chaddock, J. A. & Acharya, K. R. Botulinum and tetanus neurotoxins: structure, function and therapeutic utility. *Trends Biochem. Sci.* **27**, 552–558 (2002).

37. Rummel, A. Two Feet on the Membrane: Uptake of Clostridial Neurotoxins. in *Uptake and Trafficking of Protein Toxins* (ed. Barth, H.) vol. 406 1–37 (Springer International Publishing, 2016).

38. Peck, M. *et al.* Historical Perspectives and Guidelines for Botulinum Neurotoxin Subtype Nomenclature. *Toxins* **9**, 38 (2017).

39. Colasante, C. *et al.* Botulinum Neurotoxin Type A is Internalized and Translocated from Small Synaptic Vesicles at the Neuromuscular Junction. *Mol. Neurobiol.* **48**, 120–127 (2013).

40. Berntsson, R. P.-A., Peng, L., Dong, M. & Stenmark, P. Structure of dual receptor binding to botulinum neurotoxin B. *Nat. Commun.* **4**, 2058 (2013).

41. Hubrecht & Carter. The 3Rs and Humane Experimental Technique: Implementing Change. *Animals* **9**, 754 (2019).

42. Russell, W. M. S. & Burch, R. L. *The Principles of Humane Experimental Technique.* (Methuen & Co Ltd., 1959).

43. Adler, S. *et al.* The Current Scientific and Legal Status of Alternative Methods to the LD50 Test for Botulinum Neurotoxin Potency Testing: The Report and Recommendations of a Zebet Expert Meeting. *Altern. Lab. Anim.* **38**, 315–330 (2010).

44. Barr, J. R. *et al.* Botulinum Neurotoxin Detection and Differentiation by Mass Spectrometry. *Emerg. Infect. Dis.* **11**, 1578–1583 (2005).

45. Pellett, S. Progress in Cell Based Assays for Botulinum Neurotoxin Detection. in *Botulinum Neurotoxins* (eds. Rummel, A. & Binz, T.) vol. 364 257–285 (Springer Berlin Heidelberg, 2012).

46. Strotmeier, J., Willjes, G., Binz, T. & Rummel, A. Human synaptotagmin-II is not a high affinity receptor for botulinum neurotoxin B and G: increased therapeutic dosage and immunogenicity. *FEBS Lett.* **586**, 310–313 (2012).

47. Wictome, M. *et al.* Development of an In Vitro Bioassay for Clostridium botulinum Type B Neurotoxin in Foods That Is More Sensitive than the Mouse Bioassay. *Appl. Environ. Microbiol.* **65**, 3787–3792 (1999).

48. Whitemarsh, R. C. M. *et al.* Novel Application of Human Neurons Derived from Induced Pluripotent Stem Cells for Highly Sensitive Botulinum Neurotoxin Detection. *Toxicol. Sci.* **126**, 426–435 (2012).

49. Bigalke, H. & Rummel, A. Botulinum Neurotoxins: Qualitative and Quantitative Analysis Using the Mouse Phrenic Nerve Hemidiaphragm Assay (MPN). *Toxins* **7**, 4895–4905 (2015).

50. Pellett, S. *et al.* Human Induced Pluripotent Stem Cell Derived Neuronal Cells Cultured on Chemically-Defined Hydrogels for Sensitive In Vitro Detection of Botulinum Neurotoxin. *Sci. Rep.* **5**, 14566 (2015).

51. Pellett, S., Tepp, W. H., Johnson, E. A. & Sesardic, D. Assessment of ELISA as endpoint in neuronal cell-based assay for BoNT detection using hiPSC derived neurons. *J. Pharmacol. Toxicol. Methods* **88**, 1–6 (2017).

52. Pellett, S., Tepp, W. H., Clancy, C. M., Borodic, G. E. & Johnson, E. A. A neuronal cell-based botulinum neurotoxin assay for highly sensitive and specific detection of neutralizing serum antibodies. *FEBS Lett.* **581**, 4803–4808 (2007).

53. Basavanna, U., Muruvanda, T., Brown, E. W. & Sharma, S. K. Development of a Cell-Based Functional Assay for the Detection of *Clostridium botulinum* Neurotoxin Types A and E. *Int. J. Microbiol.* **2013**, 1–7 (2013).

54. Hong, W. S., Young, E. W. K., Tepp, W. H., Johnson, E. A. & Beebe, D. J. A Microscale Neuron and Schwann Cell Coculture Model for Increasing Detection Sensitivity of Botulinum Neurotoxin Type A. *Toxicol. Sci.* **134**, 64–72 (2013).

55. Rust, A. *et al.* A Cell Line for Detection of Botulinum Neurotoxin Type B. *Front. Pharmacol.* **8**, 796 (2017).

56. Jenkinson, S. P. *et al.* Embryonic Stem Cell-Derived Neurons Grown on Multi-Electrode Arrays as a Novel In vitro Bioassay for the Detection of Clostridium botulinum Neurotoxins. *Front. Pharmacol.* **8**, (2017).

57. Fernández-Salas, E. *et al.* Botulinum Neurotoxin Serotype a Specific Cell-Based Potency Assay to Replace the Mouse Bioassay. *PLoS ONE* **7**, e49516 (2012).

58. Ferreira, J. L., Eliasberg, S. J., Edmonds, P. & Harrison, M. A. Comparison of the Mouse Bioassay and Enzyme-Linked Immunosorbent Assay Procedures for the Detection of Type A Botulinal Toxin in Food. *J. Food Prot.* **67**, 203–206 (2004).

59. Hallis, B., James, B. A. & Shone, C. C. Development of novel assays for botulinum type A and B neurotoxins based on their endopeptidase activities. *J. Clin. Microbiol.* **34**, 1934–1938 (1996).

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

61. Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154–156 (1981).

62. Waddington, S. N., Privolizzi, R., Karda, R. & O'Neill, H. C. A Broad Overview and Review of CRISPR-Cas Technology and Stem Cells. *Curr. Stem Cell Rep.* **2**, 9–20 (2016).

63. Thomson, J. A. *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145–1147 (1998).

64. Takahashi, K. & Yamanaka, S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* **126**, 663–676 (2006).

65. Hockemeyer, D. & Jaenisch, R. Induced Pluripotent Stem Cells Meet Genome Editing. *Cell Stem Cell* **18**, 573–586 (2016).

66. Wigler, M. *et al.* Transfer of purified herpes virus thymidine kinase gene to cultured mouse cells. *Cell* **11**, 223–232 (1977).

67. Folger, K. R., Wong, E. A., Wahl, G. & Capecchi, M. R. Patterns of Integration of DNA Microinjected into Cultured Mammalian Cells: Evidence for Homologous Recombination Between Injected Plasmid DNA Molecules. *MOL CELL BIOL* **2**, 16 (1982).

68. Smithies, O., Gregg, R. G., Boggst, S. S., Koralewski, M. A. & Kucherlapati, R. S. Insertion of DNA sequences into the human chromosomal p-globin locus by homologous recombination. *Nature* 5 (1985).

69. Jinek, M. *et al.* A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816–821 (2012).

70. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).

71. Sadelain, M., Papapetrou, E. P. & Bushman, F. D. Safe harbours for the integration of new DNA in the human genome. *Nat. Rev. Cancer* **12**, 51–58 (2012).

72. Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84–88 (2016).

73. Mali, P. *et al.* RNA-Guided Human Genome Engineering via Cas9. *Science* **339**, 823–826 (2013).

74. Oceguera-Yanez, F. *et al.* Engineering the AAVS1 locus for consistent and scalable transgene expression in human iPSCs and their differentiated derivatives. *Methods* **101**, 43–55 (2016).

75. Markova, S. V., Larionova, M. D. & Vysotski, E. S. Shining Light on the Secreted Luciferases of Marine Copepods: Current Knowledge and Applications. *Photochem. Photobiol.* **95**, 705–721 (2019).

76. Shastry, P., Basu, A. & Rajadhyaksha, M. S. Neuroblastoma Cell Lines-A Versatile in Vztro Model in Neurobiology. *Int. J. Neurosci.* **108**, 109–126 (2001).

77. Goodall, A. R., Danks, K., Walker, J. H., Ball, S. G. & Vaughan, P. F. T. Occurrence of Two Types of Secretory Vesicles in the Human Neuroblastoma SH-SY5Y. *J. Neurochem.* **68**, 1542–1552 (2002).

78. Sobota, J. A., Mohler, W. A., Cowan, A. E., Eipper, B. A. & Mains, R. E. Dynamics of peptidergic secretory granule transport are regulated by neuronal stimulation. *BMC Neurosci.* **11**, 32 (2010).

79. Prado, V. F. & Prado, M. A. M. Signals Involved in Targeting Membrane Proteins to Synaptic Vesicles. *Cell. Mol. Neurobiol.* 13 (2002).

80. Winkler, H., Sietzen, M. & Schober, M. The Life Cycle of Catecholamine-storing Vesicles. *Ann. N. Y. Acad. Sci.* **493**, 3–19 (1987).

81. Bonifacino, J. S. & Glick, B. S. The Mechanisms of Vesicle Budding and Fusion. *Cell* **116**, 153–166 (2004).

82. Kelly, R. B. Secretory granule and synaptic vesicle formation. *Curr. Opin. Cell Biol.* **3**, 654–660 (1991).

83. Coughenour, H. D., Spaulding, R. S. & Thompson, C. M. The synaptic vesicle proteome: A comparative study in membrane protein identification. *PROTEOMICS* **4**, 3141–3155 (2004).

84. Fernández-Chacón, R. & Südhof, T. C. GENETICS OF SYNAPTIC VESICLE FUNCTION: Toward the Complete Functional Anatomy of an Organelle. *Annu. Rev. Physiol.* **61**, 753–776 (1999).

85. Nunn, C. M. *et al.* Crystal structure of tobacco etch virus protease shows the protein C terminus bound within the active site. *J. Mol. Biol.* **350**, 145–155 (2005).

86. Verhaegent, M. & Christopoulos, T. K. Recombinant Gaussia luciferase. Overexpression, purification, and analytical application of a bioluminescent reporter for DNA hybridization. *Anal. Chem.* **74**, 4378–4385 (2002).

87. Pennuto, M., Bonanomi, D., Benfenati, F. & Valtorta, F. Synaptophysin I Controls the Targeting of VAMP2/Synaptobrevin II to Synaptic Vesicles. *Mol. Biol. Cell* **14**, 4909–4919 (2003).

88. Kroczek, R. A. Southern and northern analysis. *J. Chromatogr.* **618**, 133–145 (1993).

89. Southern, E. M. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**, 503–517 (1975).

90. Yu, D. *et al.* Cyclic Digestion and Ligation-Mediated PCR Used for Flanking Sequence Walking. *Sci. Rep.* **10**, 3434 (2020).

91. Aniento, F. & Gruenberg, J. Subcellular Fractionation of Tissue Culture Cells. *Curr. Protoc. Immunol.* **57**, (2003).

92. Pasquali, C., Fialka, I. & Huber, L. A. Subcellular fractionation, electromigration analysis and mapping of organelles. *J. Chromatogr. B. Biomed. Sci. App.* **722**, 89–102 (1999).

93. Bolte, S. & Cordelières, F. P. A guided tour into subcellular colocalization analysis in light microscopy. *J. Microsc.* **224**, 213–232 (2006).

94. Dunn, K. W., Kamocka, M. M. & McDonald, J. H. A practical guide to evaluating colocalization in biological microscopy. *Am. J. Physiol.-Cell Physiol.* **300**, C723–C742 (2011).

95. Jensen, E. Technical review: Colocalization of antibodies using confocal microscopy: Colocalization of Antibodies. *Anat. Rec.* **297**, 183–187 (2014).

96. Zucker, R. Photorelease Techniques for Raising or Lowering Intracellular Ca2+. *Methods Cell Biol.* **40**, 31–63 (1994).

97. pUNO vectors. *InvivoGen* https://www.invivogen.com/puno (2016).

98. VAMP2 (NM_014232) Human Tagged ORF Clone – RC207533 | OriGene. https://www.origene.com/catalog/cdna-clones/expression-plasmids/rc207533/vamp2-nm_014232-human-tagged-orf-clone.

99. Schenke, M., Schjeide, B.-M., Püschel, G. P. & Seeger, B. Analysis of Motor Neurons Differentiated from Human Induced Pluripotent Stem Cells for the Use in Cell-Based Botulinum Neurotoxin Activity Assays. *Toxins* **12**, 276 (2020).

100. O'Malley, R. C., Alonso, J. M., Kim, C. J., Leisse, T. J. & Ecker, J. R. An adapter ligation-mediated PCR method for high-throughput mapping of T-DNA inserts in the Arabidopsis genome. *Nat. Protoc.* **2**, 2910–2917 (2007).

101. Vandesompele, J., Preter, K. D., Roy, N. V. & Paepe, A. D. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. 12.

102. Yu, Z., Huang, Z. & Lung, M. Subcellular Fractionation of Cultured Human Cell Lines. *BIO-Protoc.* **3**, (2013).

103. Berg, J. M., Tymoczko, J. L. & Stryer, L. Section 4.1, The Purification of Proteins Is an Essential First Step in Understanding Their Function. in *Biochemistry* (W H Freeman, 2002).

104. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254 (1976).

105. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).

106. Manders, E. M., Stap, J., Brakenhoff, G. J., van Driel, R. & Aten, J. A. Dynamics of three-dimensional replication patterns during the S-phase, analysed by double labelling of DNA and confocal microscopy. *J. Cell Sci.* **103 (Pt 3)**, 857–862 (1992).

107. Norrman, K. *et al.* Quantitative comparison of constitutive promoters in human ES cells. *PloS One* **5**, e12413 (2010).

108. Qin, J. Y. *et al.* Systematic Comparison of Constitutive Promoters and the Doxycycline-Inducible Promoter. *PLoS ONE* **5**, e10611 (2010).

109. Teschendorf, C., Warrington, K. H., Siemann, D. W. & Muzyczka, N. Comparison of the EF-1 alpha and the CMV promoter for engineering stable tumor cell lines using recombinant adeno-associated virus. *Anticancer Res.* **22**, 3325–3330 (2002).

110. Nakano, A. & Luini, A. Passage through the Golgi. *Curr. Opin. Cell Biol.* **22**, 471–478 (2010).

111. Kroehne, V. *et al.* Primary Spinal OPC Culture System from Adult Zebrafish to Study Oligodendrocyte Differentiation In Vitro. *Front. Cell. Neurosci.* **11**, 284 (2017).

112. Ehlken, H. *et al.* Hepatocyte IKK2 Protects Mdr2−/− Mice from Chronic Liver Failure. *PLoS ONE* **6**, e25942 (2011).

113. AcTEV™ Protease. https://www.thermofisher.com/order/catalog/product/12575015.

114. Renden, R. & von Gersdorff, H. Synaptic Vesicle Endocytosis at a CNS Nerve Terminal: Faster Kinetics at Physiological Temperatures and Increased Endocytotic Capacity During Maturation. *J. Neurophysiol.* **98**, 3349–3359 (2007).

115. Verhage, M. *et al.* Differential release of amino acids, neuropeptides, and catecholamines from isolated nerve terminals. *Neuron* **6**, 517–524 (1991).

116. Zhang, Z. *et al.* Release mode of large and small dense-core vesicles specified by different synaptotagmin isoforms in PC12 cells. *Mol. Biol. Cell* **22**, 2324–2336 (2011).

117. Qin, L. *et al.* Novel LMW glutenin subunit genes from wild emmer wheat (Triticum turgidum ssp. dicoccoides) in relation to Glu-3 evolution. *Dev. Genes Evol.* **225**, 31–37 (2015).

118. pEF-BOSEX sequence Shigekazu Nagata Lab. | Department of Medical Chemistry Graduate School of Medicine, Kyoto University. http://www2.mfour.med.kyoto-u.ac.jp/nagata/english/research/pef-bosex.html.

119. Kapust, R. B. & Waugh, D. S. Controlled Intracellular Processing of Fusion Proteins by TEV Protease. *Protein Expr. Purif.* **19**, 312–318 (2000).

120. Green, M. R. & Sambrook, J. Polymerase Chain Reaction (PCR) Amplification of GC-Rich Templates. *Cold Spring Harb. Protoc.* **2019**, pdb.prot095141 (2019).

121. Yang, H. *et al.* Methods Favoring Homology-Directed Repair Choice in Response to CRISPR/Cas9 Induced-Double Strand Breaks. *Int. J. Mol. Sci.* **21**, 6461 (2020).

122. Carter, M. & Shieh, J. C. Cell Culture Techniques. in *Guide to Research Techniques in Neuroscience* 281–296 (Elsevier, 2010). doi:10.1016/B978-0-12-374849-2.00013-6.

123. Simple Nucleotide Polymorphisms (dbSNP 151) Found in >= 1% of Samples. http://genome-euro.ucsc.edu/cgi-bin/hgc?hgsid=240860933_BCpTrIu8xza7KSuhtFiZKkq3wTOu&c=chr19&l=55115457&r=55115468&o=55115457&t=55115458&g=snp151Common&i=rs667451.

124. Hornstra, I. K. & Yang, T. P. In Vivo Footprinting and Genomic Sequencing by Ligation-Mediated PCR. *Anal. Biochem.* **213**, 179–193 (1993).

125. Arya, M. *et al.* Basic principles of real-time quantitative PCR. *Expert Rev. Mol. Diagn.* **5**, 209–219 (2005).

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

127. Tichy, E. D. Mechanisms maintaining genomic integrity in embryonic stem cells and induced pluripotent stem cells. *Exp. Biol. Med.* **236**, 987–996 (2011).

128. Maruyama, T. *et al.* Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat. Biotechnol.* **33**, 538–542 (2015).

129. Song, F. & Stieger, K. Optimizing the DNA Donor Template for Homology-Directed Repair of Double-Strand Breaks. *Mol. Ther. - Nucleic Acids* **7**, 53–60 (2017).

130. Zhang, J.-H., Adikaram, P., Pandey, M., Genis, A. & Simonds, W. F. Optimization of genome editing through CRISPR-Cas9 engineering. *Bioengineered* **7**, 166–174 (2016).

131. Liu, X. *et al.* Sequence features associated with the cleavage efficiency of CRISPR/Cas9 system. *Sci. Rep.* **6**, 19675 (2016).

132. Kleinstiver, B. P. *et al.* High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490–495 (2016).

133. Huang, J., Wang, Y. & Zhao, J. CRISPR editing in biological and biomedical investigation. *J. Cell. Physiol.* **233**, 3875–3891 (2018).

134. Slembrouck, D., Partoens, P., Annaert, W. & De Potter, W. P. Secretory Vesicle-Specific Antibodies in the Confocal Study of Exo–Endocytosis Dynamics. *Methods* **18**, 465–471 (1999).

135. Fritschy, J.-M. Is my antibody-staining specific? How to deal with pitfalls of immunohistochemistry. *Eur. J. Neurosci.* **28**, 2365–2370 (2008).

136. Lorincz, A. & Nusser, Z. Specificity of Immunoreactions: The Importance of Testing Specificity in Each Method. *J. Neurosci.* **28**, 9083–9086 (2008).

137. Fischer-Colbrie, R., Laslop, A. & Kirchmair, R. Secretogranin II: Molecular properties, regulation of biosynthesis and processing to the neuropeptide secretoneurin. *Prog. Neurobiol.* **46**, 49–70 (1995).

138. Munro, S. & Nichols, B. J. The GRIP domain – a novel Golgi-targeting domain found in several coiled-coil proteins. *Curr. Biol.* **9**, 377–380 (1999).

139. Yoshino, A. A role for GRIP domain proteins and/or their ligands in structure and function of the trans Golgi network. *J. Cell Sci.* **116**, 4441–4454 (2003).

140. Golgin-97 Antibody (A-21270). https://www.thermofisher.com/antibody/product/Golgin-97-Antibody-clone-CDF4-Monoclonal/A-21270.

141. Jin, L.-W., Maezawa, I., Vincent, I. & Bird, T. Intracellular Accumulation of Amyloidogenic Fragments of Amyloid-β Precursor Protein in Neurons with Niemann-Pick Type C Defects Is Associated with Endosomal Abnormalities. *Am. J. Pathol.* **164**, 975–985 (2004).

142. Dieriks, B. Spatiotemporal behavior of nuclear cyclophilin B indicates a role in RNA transcription. *Int. J. Mol. Med.* (2012) doi:10.3892/ijmm.2012.937.

143. Dvoriantchikova, G., Ivanov, D., Pestova, A. & Shestopalov, V. Molecular characterization of pannexins in the lens. *Mol. Vis.* **12**, 1417–1426 (2006).

144. Immunofluorescence Troubleshooting- Protocol Tips | StressMarq. *StressMarq Biosciences Inc.* https://www.stressmarq.com/support/technical-support/troubleshooting/immunofluorescence-troubleshooting/ (2015).

145. Troubleshooting Immunofluorescence. https://www.hycultbiotech.com/media/wysiwyg/Troubleshooting_Immunofluorescence_1.pdf (2010).

146. Ihida, K., Tsuyama, S., Kashio, N. & Murata, F. Subcompartment sugar residues of gastric surface mucous cells studied with labeled lectins. *Histochemistry* **95**, 329–335 (1991).

147. Pavelka, M. & Ellinger, A. Localization of binding sites for concanavalin A, Ricinus communis I and Helix pomatia lectin in the Golgi apparatus of rat small intestinal absorptive cells. *J. Histochem. Cytochem.* **33**, 905–914 (1985).

148. Anti-Islet 1 antibody [1H9] KO Tested (ab86472) | Abcam. https://www.abcam.com/Islet-1-antibody-1H9-ab86472.html.

149. Anti-Islet 1 antibody [1B1] KO Tested (ab86501) | Abcam. https://www.abcam.com/islet-1-antibody-1b1-ab86501.html.

150. UniProt ISL1. https://www.uniprot.org/uniprot/P61371.

151. UniProt PTPRCAP. https://www.uniprot.org/uniprot/Q14761.

152. UniProt Transmembrane protein 231. https://www.uniprot.org/uniprot/B7Z5I3.

153. Schiffer, D., Cordera, S., Giordana, M. T., Attanasio, A. & Pezzulo, T. Synaptic vesicle proteins, synaptophysin and chromogranin A in amyotrophic lateral sclerosis. *J. Neurol. Sci.* **129**, 68–74 (1995).

154. Schrott-Fischer, A. *et al.* Chromogranin peptides in amyotrophic lateral sclerosis. *Regul. Pept.* **152**, 13–21 (2009).

155. Lahr, G., Mayerhofer, A., Bergmann, M., Takiyyuddin, M. A. & Gratzl, M. Chromogranin A in neurons of the rat cerebellum and spinal cord: quantification and sites of expression. *J. Histochem. Cytochem.* **40**, 993–999 (1992).

156. Smallcombe, A. Multicolor Imaging: The Important Question of Co-Localization. *BioTechniques* **30**, 1240–1246 (2001).

157. Dandoy-Dron, F., Griffond, B., Mishal, Z., Tovey, M. G. & Dron, M. Scrg1, a novel protein of the CNS is targeted to the large dense-core vesicles in neuronal cells. *Eur. J. Neurosci.* **18**, 2449–2459 (2003).

158. Hudson, A. W. Targeting of the 'insulin-responsive' glucose transporter (GLUT4) to the regulated secretory pathway in PC12 cells [published erratum appears in J Cell Biol 1993 Sep;122(5):following 1143]. *J. Cell Biol.* **122**, 579–588 (1993).

159. Monterrat, C., Grise, F., Benassy, M. N., Hémar, A. & Lang, J. The calcium-sensing protein synaptotagmin 7 is expressed on different endosomal compartments in endocrine, neuroendocrine cells or neurons but not on large dense core vesicles. *Histochem. Cell Biol.* **127**, 625–632 (2007).

160. Pan, J.-A., Fan, Y., Gandhirajan, R. K., Madesh, M. & Zong, W.-X. Hyperactivation of the Mammalian Degenerin MDEG Promotes Caspase-8 Activation and Apoptosis. *J. Biol. Chem.* **288**, 2952–2963 (2013).

161. Johnston, J. A., Ward, C. L. & Kopito, R. R. Aggresomes: A Cellular Response to Misfolded Proteins. *J. Cell Biol.* **143**, 1883–1898 (1998).

162. Perlman, D. & Halvorson, H. O. A putative signal peptidase recognition site and sequence in eukaryotic and prokaryotic signal peptides. *J. Mol. Biol.* **167**, 391–409 (1983).

163. Schenke, M. Development of an assay for the evaluation of Botulinum neurotoxin activity based on transgenic human stem cells differentiated to motor neurons. (University of Veterinary Medicine Hannover, 2020).

164. Hartjes, K. A. *et al.* Selection Via Pluripotency-Related Transcriptional Screen Minimizes the Influence of Somatic Origin on iPSC Differentiation Propensity: Transcriptional Screen Minimizes iPSC Variability. *STEM CELLS* **32**, 2350–2359 (2014).

165. Rosner, M. & Hengstschläger, M. Intercellular protein expression variability as a feature of stem cell pluripotency. *Amino Acids* **45**, 1315–1317 (2013).

166. Yu, J. *et al.* Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. *Science* **318**, 1917–1920 (2007).

167. Solomon, E. *et al.* Global transcriptome profile of the developmental principles of in vitro iPSC-to-motor neuron differentiation. *BMC Mol. Cell Biol.* **22**, 13 (2021).

168. Cho, H.-H. *et al.* Isl1 Directly Controls a Cholinergic Neuronal Identity in the Developing Forebrain and Spinal Cord by Forming Cell Type-Specific Complexes. *PLoS Genet.* **10**, e1004280 (2014).

169. Südhof, T. C. Composition of Synaptic Vesicles. in *Basic Neurochemistry: Molecular, Cellular and Medical Aspects.* (eds. Siegel, G., Agranoff, B. & Albers, R.) (Lippincott-Raven, 1999).

170. 21331 - DMEM/F-12, no glutamine, no HEPES - DE. https://www.thermofisher.com/de/de/home/technical-resources/media-formulation.329.html.

171. NEUROBASAL™ Medium (1X) liquid - DE. https://www.thermofisher.com/de/de/home/technical-resources/media-formulation.251.html.

172. Yudkoff, M. *et al.* Brain Amino Acid Requirements and Toxicity: The Example of Leucine. *J. Nutr.* **135**, 1531S-1538S (2005).

173. Velasco, I., Velasco-Velázquez, M. A., Salazar, P., Lajud, N. & Tapia, R. Influence of serum-free medium on the expression of glutamate transporters and the susceptibility to glutamate toxicity in cultured cortical neurons: Glutamate Toxicity in Serum-Free Medium. *J. Neurosci. Res.* **71**, 811–818 (2003).

174. Chinopoulos, C., Gerencser, A. A., Doczi, J., Fiskum, G. & Adam-Vizi, V. Inhibition of glutamate-induced delayed calcium deregulation by 2-APB and La3+ in cultured cortical neurones. *J. Neurochem.* **91**, 471–483 (2004).

175. Bardy, C. *et al.* Neuronal medium that supports basic synaptic functions and activity of human neurons in vitro. *Proc. Natl. Acad. Sci.* **112**, E2725–E2734 (2015).

176. Oland, L. A. & Oberlander, H. Factors that influence the development of cultured neurons from the brain of the mothManduca sexta. *Vitro Cell. Dev. Biol. - Anim.* **30**, 709–716 (1994).

177. Yang, M., Donaldson, A. E., Jiang, Y. & Iacovitti, L. Factors influencing the differentiation of dopaminergic traits in transplanted neural stem cells. *Cell. Mol. Neurobiol.* **23**, 851–864 (2003).

178.   Biffi, E., Regalia, G., Menegon, A., Ferrigno, G. & Pedrocchi, A. The Influence of Neuronal Density and Maturation on Network Activity of Hippocampal Cell Cultures: A Methodological Study. *PLoS ONE* **8**, e83899 (2013).

179.   Park, Y.-S. *et al.* Activity-Dependent Potentiation of Large Dense-Core Vesicle Release Modulated by Mitogen-Activated Protein Kinase/Extracellularly Regulated Kinase Signaling. *Endocrinology* **147**, 1349–1356 (2006).

180.   Hoogstraaten, R. I., van Keimpema, L., Toonen, R. F. & Verhage, M. Tetanus insensitive VAMP2 differentially restores synaptic and dense core vesicle fusion in tetanus neurotoxin treated neurons. *Sci. Rep.* **10**, 10913 (2020).

181.   Chen, Y. A. & Scheller, R. H. SNARE-mediated membrane fusion. *Nat. Rev. Mol. Cell Biol.* **2**, 98–106 (2001).

182.   Gümürdü, A. *et al.* MicroRNA exocytosis by large dense-core vesicle fusion. *Sci. Rep.* **7**, 45661 (2017).

183.   Pakala, R. S., Brown, K. N. & Preuss, C. V. Cholinergic Medications. in *StatPearls* (StatPearls Publishing, 2020).

184.   Müller, W. & Connor, J. Ca2+ signalling in postsynaptic dendrites and spines of mammalian neurons in brain slice. *J. Physiol.-Paris* **86**, 57–66 (1992).

185.   Weng, F. J. *et al.* Carbachol excites sublaterodorsal nucleus neurons projecting to the spinal cord: Muscarinic activation of pontine atonia neurons. *J. Physiol.* **592**, 1601–1617 (2014).

186.   Mita, A., Yoshida, M. & Nagayama, T. Nitric oxide modulates a swimmeret beating rhythm in the crayfish. *J. Exp. Biol.* **217**, 4423–4431 (2014).

6 Appendix

6.1 List of consumables used in project

Item	Company	Product number
15/50 ml CELLSTAR® Polypropylene tubes	Greiner	188261 / 227270
2ml Cryogenic vials	Carl Roth	E309.1
Accutase	Sigma	A6964-500ML
Agarose NEEO Ultra-Quality	Carl Roth	2267.4
Agarose-out DNA Purification Kit	roboklon	E3540
Ampicillin	Gibco	11593027
Anti-Digoxigenin-AP Fab fragments	Roche	11 093 274 910
Ascorbic Acid	Sigma	A4544
B-27® Supplement (50X), minus vitamin A	Thermo Fisher Scientific	12587010
BDNF	Peprotech	450-02
Brand® Pipette tip, filter tip 0.5-20µl	Sigma	Z740037
Carbomoylcholine chloride	Sigma	C4382
CDP-Star	Roche	11685627001
Compound E	Bertin Pharma	15579
Cell culture 96/24/6-well plates	Sarstedt	83.3924 / 83.3922 / 83.3920
Cell culture clear bottom microplate (96-wells)	Greiner	655088
Cell culture white microplate (96-wells)	Greiner	655073
Cell scraper	Sarstedt	83.3952
CHIR 99021	axon medchem	Axon 1386
Chromatography-Papers Whatman®	Carl Roth	7604.1
ClipTip™ pipette tips (200µl)	Thermo Fisher Scientific	94410313
CloneR™	Stemcell Technologies	5888
CNTF	Peprotech	450-13
Coelenterazine	Carl Roth	4094.3
Collagenase, Type IV, powder	Thermo Fisher Scientific	17104019
60mm non TC-Treated Culture Dish	Corning	430589
CSPD	Roche	11 655 884
DAPT	Hycultec	HY-13027
dbcAMP	Hycultec	HY-B0764
DIG Wash and Block Buffer Set	Roche	11585762001
Dimethy Sulfoxide for cell culture	PanReac AmpliChem	A3672

DMEM/F12 + HEPES	Thermo Fisher Scientific	31330095
DMEM/F12, no glutamine	Thermo Fisher Scientific	21331046
DMH1	Bertin Pharma	16679
DNA Away	Carl Roth	7010
DNeasy® Blood & Tissue Kit	QIAGEN	69504
dNTP Set, 100 mM Solutions	Thermo Fisher Scientific	R0182
Dorsomorphin (Compound C)	abcam	ab120843
DreamTaq DNA Polymerase (5 U/µL)	Thermo Fisher Scientific	EP0702
EGTA	Carl Roth	3054.3
Electroporation cuvette 2mm	Bulldog Bio	12358-2U
Ethidium Bromide	Thermo Fisher Scientific	15585011
Ethylendiamine tetraacetic acid	Carl Roth	8040.3
FastDigest restriction enzymes	Thermo Fisher Scientific	various
FBS Standard	PAN Biotech	P30-3306
GDNF	Peprotech	450-10
GeneJET Plasmid Maxiprep Kit	Thermo Fisher Scientific	K0492
GeneJET Plasmid Miniprep Kit	Thermo Fisher Scientific	K0503
Hard-Shell® low profile thin wall 96-well skirted PCR plates	Bio-Rad	HSP9601
HEPES PUFFERAN®	Carl Roth	9105.3
Knockout DMEM	Thermo Fisher Scientific	10829018
Maleic Acid	Sigma	M0375
MassRuler DNA ready-to-use	Thermo Fisher Scientific	SM0403
Matrigel (High Concentration, growth factor reduced)	Corning	354263
Maxima SYBR Green/ROX qPCR Master Mix (2X)	Thermo Fisher Scientific	K0222
Microscope Cover Glasses	Paul Marienfeld GmbH & Co. KG	111520
Microscope slides	Carl Roth	H868
Microseal® 'B' adhesive seals	Bio-Rad	MSB1001
Multiply® - Pro cup 0.2ml, PP	Sarstedt	72.737.002
Multiply® - Pro cup 0.5ml, PP	Sarstedt	72.735.002
N-2 Supplement (100X)	Thermo Fisher Scientific	17502048
Non-essential amino acids	Merck/Millipore	K0293

Neurobasal® Medium	Thermo Fisher Scientific	21103049
Oligo (dt) 18 [500ng/µl]	Thermo Fisher Scientific	SO131
Parafilm® M	Sigma	P7793
Paraformaldehyde	Roth	335.2
Passive Lysis Buffer (5x)	Promega	E194A
Pasteur pipettes	Carl Roth	4518.1
P-BoNT/A1	miprolab	3101
Penicillin-Streptomycin	PAN Biotech	P06-07100
Phusion™ High-Fidelity DNA Polymerase (2 U/µL)	Thermo Fisher Scientific	F530S
Pipette tip 10µl	Carl Roth	KP26.1
Pipette tip 1000µl	Sarstedt	70.3050.020
Pipette tip 200µl	Sarstedt	70.760.002
Purmorphamine	Stemcell Technologies	72202
ReliaPrep RNA Tissue Miniprep System	Promega	Z6010
Retinoic Acid	Sigma	R2625-50MG
RevertAid RT Reverse Transcription Kit	Thermo Fisher Scientific	K1691
ROCK-Inhibitor Y-27632	Bertin Pharma	T1725
RPMI Medium 1640 + GLuctaMAX™	Gibco	61870036
SAG	Hycultec	HY-12848
SB431542	Stemcell Technologies	72232
Sodium chloride	Carl Roth	3957.2
Sodium hydroxide	Carl Roth	6771.1
StemMACS™ iPS-Brew XF, human	Miltenyi	130-104-368
T4 DNA Ligase (5 U/µL)	Thermo Fisher Scientific	EL0011
TRIS PUFFERAN®	Carl Roth	5429-2
Triton X 100	Carl Roth	3051.2
TurboFect Transfection Reagent	Thermo Fisher Scientific	R0533
VECTASHIELD® Antifade Mounting Medium	Vector Laboratories	H-1000-10
Primary antibodies		
Anti-Gaussia luciferase (rabbit)	NEB	E8023
Anti-Gaussia luciferase (SC) (rabbit)	Santa Cruz	IT-000-014
Anti-Golgin-97 Monoclonal Antibody (CDF4) (mouse)	Thermo Fisher Scientific	A-21270
Anti-Lectin HPA Alexa Fluor R 647	Thermo Fisher Scientific	L32454
Anti-GM130 - Clone 35/GM130 (mouse)	BD Biosciences	610823

Anti-Chr-A Antibody (C-20)	Santa Cruz	sc-1488
Anti-Secretogranin II (rabbit)	Thermo Fisher Scientific	PA1-10838
Anti-Chromogranin A (CGA/414)) (mouse)	Novus Biologicals	NBP2-29428
Anti-Chromogranin C Polyclonal Antibody (SgII) (mouse)	Thermo Fisher Scientific	pa1-10838
Anti-Synaptophysin (7H12) Mouse mAb	Cell Signaling Technology	12270
Anti-GAPDH (rabbit)	Santa Cruz	Sc-25778
Anti-Islet 1 antibody [1B1] (mouse)	Abcam	ab86501
Anti-Islet 1 antibody [1H9] (mouse)	Abcam	ab86472
Secondary Alexa-fluor labelled antibodies		
Anti-mouse IgG (H+L) (Alexa 488)	Invitrogen	A11001
Anti-rabbit IgG (H+L) (Alexa 488)	Invitrogen	A32731
Anti-rabbit IgG (H+L) (Alexa 568)	Invitrogen	A11036
Anti-goat IgG (H+L) (Alexa 488)	Invitrogen	A11055

6.2   List of equipment used in project

Item	Company
Automatic pipette helper, pepetus® akku	Hirschamnn Laborgeräte Eberstadt
Automatic pipette helper, PipetAid XP	Drummond
Automatic pipette, Finnpipette	ThermoLabSystem
Automatic pipette, Gilson (P10, P20, P200, P1000)	Eppendorf AG
Bensen burner Gasprofi 1	Wartewig-Labortechnik
C1000™ Thermal Cylcer CFX96™ Real-Time System	Bio-Rad
Cell culture laminar flow hood	NuAire
ChemiDoc™ MP Imaging System	Bio-Rad
E1-ClipTip™ Equalizer multichannel pipette	Thermo Fisher Scientific
EVE™ Automatic cell counter	NanoEnTek
Fluorescence microscope DM6 B with CTR6 LED	Leica
FLUOstar OMEGA	BMG LABTECH
Gel Doc™ EZ Imager	Bio-Rad
Gel electrophoresis chamber	VWR
GENE PULSER® II	Bio-Rad
Heraeus Multifuge X1R Centrifuge	Thermo Fisher Scientific
Heraeus Pico 17 Centrifuge	Thermo Fisher Scientific
High Voltage Power Pac 300	Bio-Rad
Hybridizer HB-1000	UVP Laboratory Product
Ice machine	Ziegra
Inverted modular microscope LSM 780 T-PMT Observer.Z1	Zeiss
Laboratory Centrifuge 3K30	Sigma

Laboratory scale	Sartorius
Magnet mixer / hot plate combination MR-3001	Heidolph
Microscope camera XC30	Olympus
Microscope CKX41	Olympus
Mircowave	VWR
NanoVue Plus	GE Healthcare Bio-Sciences AB
Optima™ LE-80K Ultracentrifuge	Beckman Coulter
pH meter inoLab	WtW
Rotor 12154-H	Sigma
T3000 Thermocycler	Biometra
Thermocycler 60	bio-med
Thermometer	VWR
Vortex genie 2	Scientific Industries
Water bath ecoline O11	Lauda
Water heater coil	VWR

Software:
GraphPad Prism
Microsoft Office
Fiji (Fiji is just ImageJ)
Inkscape
SnapGene
Benchling
Zotero
Image Lab
ZEN 2012 SP5

6.3 Gaussia luciferase construct sequences

6.3.1 DNA sequences for all GLuc constructs

The following section contains the complete sequences of the sorting tag and GLuc constructs, including restriction enzyme recognition sites, which were cloned into the pAAVS1-P-MCS donor plasmid and which were used to create clones to be characterized for the MoN-Light BoNT assay. Each DNA segment is color-coded.

no tag GLuc

Complete sequence of no tag GLuc construct, including restriction enzyme recognition sites, Kozak sequence, GLuc sequence, and Ef1-HTLV promoter sequence.

```
5'ACTAGTctagtcaccaccggccccccttgatcttgtccacctggccctggatcttgctggcaaaggt
cgcacagcgttgcggcagccacttcttgagcaggtcagaacactgcacgttggcaagccctttgaggc
agccagttgtgcagtccacacacagatcgacctgtgcgatgaactgctccatgggctccaagtccttg
aacccaggaatctcaggaatgtcgacgatcgcctcgcctatgccgccctgtgcggactctttgtcgcc
ttcgtaggtgtggcagcgtcctgggatgaacttcttcatcttgggcgtgcacttgatgtgggacaggc
agatcagacagcccctggtgcagccagctttccgggcattggcttccatctctttgagcacctccagc
ggcagcttcttgccgggcaacttcccgcggtcagcatcgagatccgtggtcgcgaagttgctggccac
ggccacgatgttgaagtcttcgttgttctcggtgggcttcatggtggTTAATTAAGATCTGTAACGGC
GCAGAACAGAAAACGAAACAAAGACGTAGAGTTGAGCAAGCAGGGTCAGGCAAAGCGTGGAGAGCCGG
CTGAGTCTAGGTAGGCTCCAAGGGAGCGCCGGACAAAGGCCCGGTCTCGACCTGAGCTTTAAACTTAC
CTAGACGGCGGACGCAGTTCAGGAGGCACCACAGGCGGGAGGCGGCAGAACGCGACTCAACCGGCGTG
GATGGCGGCCTCAGGTAGGGCGGCGGGCGCGTGAAGGAGAGATGCGAGCCCCTCGAAGCTTCAGCTGT
GTTCTGGCGGCAAACCCGTTGCGAAAAAGAACGTTCACGGCGACTACTGCACTTATATACGGTTCTCC
CCCACCCTCGGGAAAAAGGCGGAGCCAGTACACGACATCACTTTCCCAGTTTACCCCGCGCCACCTTC
TCTAGGCACCGGTTCAATTGCCGACCCCTCCCCCCAACTTCTCGGGGACTGTGGGCGATGTGCGCTCT
GCCCACTGACGTCGAC-3'
```

hPOMC-GLuc

Complete sequence of hPOMC-GLuc construct, including restriction enzyme recognition sites, Kozak sequence, GLuc sequence, hPOMC AA1-26 sequence, and Ef1-HTLV promoter sequence.

```
5'ACTAGTctagtcaccaccggccccccttgatcttgtccacctggccctggatcttgctggcaaaggt
cgcacagcgttgcggcagccacttcttgagcaggtcagaacactgcacgttggcaagccctttgaggc
agccagttgtgcagtccacacacagatcgacctgtgcgatgaactgctccatgggctccaagtccttg
aacccaggaatctcaggaatgtcgacgatcgcctcgcctatgccgccctgtgcggactctttgtcgcc
ttcgtaggtgtggcagcgtcctgggatgaacttcttcatcttgggcgtgcacttgatgtgggacaggc
agatcagacagcccctggtgcagccagctttccgggcattggcttccatctctttgagcacctccagc
ggcagcttcttgccgggcaacttcccgcggtcagcatcgagatccgtggtcgcgaagttgctggccac
ggccacgatgttgaagtcttcgttgttctcggtgggcttcatgccacgcacttccatggaggcctgaa
gcagcaaggccagcaacagggccccccgagcggctgcagcacgatctcggcatggtggTTAATTAAGAT
CTGTAACGGCGCAGAACAGAAAACGAAACAAAGACGTAGAGTTGAGCAAGCAGGGTCAGGCAAAGCGT
GGAGAGCCGGCTGAGTCTAGGTAGGCTCCAAGGGAGCGCCGGACAAAGGCCCGGTCTCGACCTGAGCT
TTAAACTTACCTAGACGGCGGACGCAGTTCAGGAGGCACCACAGGCGGGAGGCGGCAGAACGCGACTC
AACCGGCGTGGATGGCGGCCTCAGGTAGGGCGGCGGGCGCGTGAAGGAGAGATGCGAGCCCCTCGAAG
CTTCAGCTGTGTTCTGGCGGCAAACCCGTTGCGAAAAAGAACGTTCACGGCGACTACTGCACTTATAT
ACGGTTCTCCCCCACCCTCGGGAAAAAGGCGGAGCCAGTACACGACATCACTTTCCCAGTTTACCCCG
CGCCACCTTCTCTAGGCACCGGTTCAATTGCCGACCCCTCCCCCCAACTTCTCGGGGACTGTGGGCGA
TGTGCGCTCTGCCCACTGACGTCGAC-3'
```

CgA-GLuc

Complete sequence of CgA-GLuc construct, including restriction enzyme recognition sites, Kozak sequence, GLuc sequence, segment of chromogranin A sequence, and Ef1-HTLV promoter sequence.

5'ACTAGTctagtcaccaccggccccccttgatcttgtccacctggccctggatcttgctggcaaaggt
cgcacagcgttgcggcagccacttcttgagcaggtcagaacactgcacgttggcaagccctttgaggc
agccagttgtgcagtccacacacagatcgacctgtgcgatgaactgctccatgggctccaagtccttg
aacccaggaatctcaggaatgtcgacgatcgcctcgcctatgccgccctgtgcggactctttgtcgcc
ttcgtaggtgtggcagcgtcctgggatgaacttcttcatcttgggcgtgcacttgatgtgggacaggc
agatcagacagcccctggtgcagccagctttccgggcattggcttccatctctttgagcacctccagc
ggcagcttcttgccgggcaacttcccgcggtcagcatcgagatccgtggtcgcgaagttgctggccac
ggccacgatgttgaagtcttcgttgttctcggtgggcttcatttggatcCGAGCTCGGTACCggcccc
gccgtagtgcctgcagctggtgggccactttctccagctctgcttcaatggccgacaggctctccagc
tcctggtcctctggtctgcggtttgcgctgccctcctcctctttcttctcctcggggtagcctcggac
ctggaggggcaggcccgcctcaaggctgtcctcccgggaggatggcctccagcctcgtcgcagctgcg
gcccagggcccctgaagccgtaggcccgggcccggaaggagagcttcatggaactgtcccggttgtcc
tcctcctcctgcccctccagccgcttctcagccgtcagctccttggccagctggtccatcttgct
ccagcgtttggagtcctcccactccttggagagccgctcctcctcctgctccagctctccgctcttcc
cacccggaagaggccttgcgggaccactgccatctcctcctcctctctttctgctgggagtgctcc
tgttctccttcccttcggggtcctgagcctcctcagccccaggcttcccagctccatccacagccag
agcctccgaccgactctcgcctttccggatctccttgtagccaaggctcgggtgggggttcagcacta
cagtggggccttcttcctcggggacagcctcctctccagcctcagcctcctcctcctcctcctcc
tcttctctctttgcctgccaccctggctctgcactcaggcccttctctctgtccaccagaccctgaga
gaggccctcactgtccccctcggcctgtgggcctgggtatttctggctggggaggctggctggaggt
gggtgttggtggcctcctcctcctcctcctcttcctccccagggggcctgattgttcccctcagccttg
gactcctgcatgggctccgggagggcctggggcctggctccgtctgtggcttcaccacttttctctgc
ctccttggaatcctctcttttctccataacatccttggatgatggctcttccaccgcctctttcagct
cggcctggctgctctggttctcaagaacctctgagagttcatcttcaaaaccgctgtgtttcttctgc
tgatgtgccctctccttggcgccttggagagcgaggtcttggagctccttcagtaaattctgatgtct
cagaatggaaaggatccgttcatctcctcggagtgtctcaaaacattcctggctgacaggcatggggc
tgggcttggaaagtgtgtcggagatgacctcaacgatgcatttcatcacctcggtatccccctttattc
atagggctgttcacagggagcgcagtgacttgcccggcgcagagcagaagagcca**gga**cagcggcgga
gcgcatggtgTTAATTAA*GATCTGTAACGGCGCAGAAC*AGAAAACGAAACAAAGACGTAGAGTTGAGC
AAGCAGGGTCAGGCAAAGCGTGGAGAGCCGGCTGAGTCTAGGTAGGCTCCAAGGGAGCGCCGGACAAA
GGCCCGGTCTCGACCTGAGCTTTAAACTTACCTAGACGGCGGACGCAGTTCAGGAGGCACCACAGGCG
GGAGGCGGCAGAACGCGACTCAACCGGCGTGGATGGCGGCCTCAGGTAGGGCGGCGGGCGCGTGAAGG
AGAGATGCGAGCCCCTCGAAGCTTCAGCTGTGTTCTGGCGGCAAACCCGTTGCGAAAAAGAACGTTCA
CGGCGACTACTGCACTTATATACGGTTCTCCCCCACCCTCGGGAAAAAGGCGGAGCCAGTACACGACA
TCACTTTCCCAGTTTACCCCGCGCCACCTTCTCTAGGCACCGGTTCAATTGCCGACCCCTCCCCCCAA
CTTCTCGGGGACTGTGGGC*GATGTGCGCTCTGCCCACTGAC*GTCGAC-3'

SgII-GLuc

Complete sequence of SgII-GLuc construct, including restriction enzyme recognition sites, Kozak sequence, GLuc sequence, segment of Secretogranin II sequence, and Ef1-HTLV promoter sequence.

5'ACTAGTctagtcaccaccggccccccttgatcttgtccacctggccctggatcttgctggcaaaggt
cgcacagcgttgcggcagccacttcttgagcaggtcagaacactgcacgttggcaagccctttgaggc
agccagttgtgcagtccacacacagatcgacctgtgcgatgaactgctccatgggctccaagtccttg

aacccaggaatctcaggaatgtcgacgatcgcctcgcctatgccgccctgtgcggactctttgtcgcc
ttcgtaggtgtggcagcgtcctgggatgaacttcttcatcttgggcgtgcacttgatgtgggacaggc
agatcagacagcccctggtgcagccagctttccgggcattggcttccatctctttgagcacctccagc
ggcagcttcttgccgggcaacttcccgcggtcagcatcgagatccgtggtcgcgaagttgctggccac
ggccacgatgttgaagtcttcgttgttctcggtgggcttcatttggatccccatattttccattgctc
tcttagcaatatgctcccttccttttctgccttttcttggttgaggtattccagcactttcattaac
agatcttcatcccagtactgcctatttggggtatcatcattcttcggggggccccacagggaaccttt t
gctcaccggggccagcttgtcagtctcctgagagctgccttgattcaaatgctctttgatggcctgct
caatttgttcctcttcctgcaggtcatcttcagatgagccttgaccaggaactcgcttcacttggttt
gaattaatgatctcagggtatttaactagcatcctggccaagtactcacctaattcttgatccttgtc
gttcaggttttcatatgccatctgtctgtttttcaacatgtggaatccaggcagctttgggaagctggt
tcgatctagatcttccagcaccataagggagccttggcagaactttctcctggttatatggattggga
aaatacgacgttttctgatttgctgcactctccatccctaaaagatttaaaatatcctcaacactgag
ccgtctggtagggcctcagtcccagcacgaccaggtgtttttagggtagccactcttggagagcatcc
tattttggaacaggtctggatggtctaagtcagcctctgagatgtcatctaggtcaacaggaaggtca
agctcccgctccggttccactgatccattcggcttctccccagttttgagcatctcaattaagtcttc
tggggggtatctgtaaattccttgagatttcaatcagctgataaatagactgagaatcaagaggtttct
caaaaagcctggtggccctttccccattttgcccattctgtaacctcccacttcctgcagcatttact
aacttttttcaaataggcaattactttggagacatcatctgagagttggtctttactctcttttccgaag
atcttcttcctggatgccaagctgccctgagcgtttcatctcatcgttgatttgttcattttttccta
tattctctttgctgtctctcacctcttcctgggtttgactctctattttctcctctactgggttccag
tcttctcccccgaccacatcttcataggcaatgttattagccttgtagatatcatcttcatcatccgt
ataaagttttttgctcctcatccatcctctcacgtttctggttgtttggtcctgtcagtttccccagct
cttggaagacagattccaatgtagcaaggctttgaggagtatattgttcctccactatttcatttgtg
cgtttaaaggggttatccctggaattctcttcatacataggagggaattgcatgtgcttaagctttct
ttctggccactgctgtgtctcataatcatcactcatgtccattggaaagttctttttctgaattcaagg
catagggcttatttttcttttggtgcagactgaggctcattttcagcctgtctcaaagcttcgagtatt
attctcatccagtcttcttcactcagtgaatccctctcgggcaagtggctttcatcgccatttttcttt
ttgctgaagggggacagagacaccttggtagggattataatctgggctgctttcttccttatgagctt
gttgtcggaggttttctatgtactccaaagccctgatcatttcaggactgggaaacttttggacattt
tccaacctgaggtctggttctttctgaagcagctggtttctctgaaatgaagctgcttcagccccaga
gatgaggaaaattaaagggataagagacagggctgctccaagccagtgggtctttgcttcagccatgg
tgTTAATTAA*GATCTGTAACGGCGCAGAAC*AGAAAACGAAACAAAGACGTAGAGTTGAGCAAGCAGGG
TCAGGCAAAGCGTGGAGAGCCGGCTGAGTCTAGGTAGGCTCCAAGGGAGCGCCGGACatgttgGGTCT
CGACCTGAGCTTTAAACTTACCTAGACGGCGGACGCAGTTCAGGAGGCACCACAGGCGGGAGGCGGCA
GAACGCGACTCAACCGGCGTGGATGGCGGCCTCAGGTAGGCGGCGGGCGCGTGAAGGAGAGATGCGA
GCCCCTCGAAGCTTCAGCTGTGTTCTGGCGGCAAACCCGTTGCGAAAAAGAACGTTCACGGCGACTAC
TGCACTTATATACGGTTCTCCCCCACCCTCGGGAAAAAGGCGGAGCCAGTACACGACATCACTTTCCC
AGTTTACCCCGCGCCACCTTCTCTAGGCACCGGTTCAATTGCCGACCCCTCCCCCAACTTCTCGGGG
ACTGTGGGC*GATGTGCGCTCTGCCCACTGAC*GTCGAC-3'

VAMP2-GLuc

Complete sequence of VAMP2-GLuc construct, including restriction enzyme recognition sites, Kozak sequence, GLuc sequence, TEV recognition sequence, VAMP2 ORF sequence, and Ef1-HTLV promoter sequence.

```
5' ACTAGTctagtcaccaccggcccccttgatcttgtccacctggccctggatcttgctggcaaaggt
cgcacagcgttgcggcagccacttcttgagcaggtcagaacactgcacgttggcaagccctttgaggc
agccagttgtgcagtccacacacagatcgacctgtgcgatgaactgctccatgggctccaagtccttg
aacccaggaatctcaggaatgtcgacgatcgcctcgcctatgccgccctgtgcggactctttgtcgcc
ttcgtaggtgtggcagcgtcctgggatgaacttcttcatcttgggcgtgcacttgatgtgggacaggc
agatcagacagcccctggtgcagccagctttccgggcattggcttccatctctttgagcacctccagc
ggcagcttcttgccgggcaacttcccgcggtcagcatcgagatccgtggtcgcgaagttgctggccac
ggccacgatgttgaagtcttcgttgttctcggtgggcttcatGAATTCTCCTTGAAAATATAAGTTTT
CCTTAAGAGCGGTACCAGTGCTGAAGTAAACTATGATGATGATGAGGATGATGGCGCAAATCACTCCC
AAGATGATCATCATCTTGAGGTTTTTCCACCAGTATTTGCGCTTGAGCTTGGCTGCGCTTGTTTCAAA
CTGGGAGGCCCCCGCCTGGAGTGCATCTGCACGGTCGTCCAGCTCCGACAGCTTCTGGTCTCGCTCCA
GGACCTTGTCCACGTTCACCCTCATGATGTCCACCACCTCATCCACCTGGGCCTGGGTCTGCTGCAGT
CTCCTGTTACTGGTGAGGTTTGGAGGGGGTGCAGGGGGACCACCCTCCCCAGCCGGGGCAGCAGGGGG
GGTCGTGGCAGCGGTAGCAGACATggtggTTAATTAAGATCTGTAACGGCGCAGAACAGAAAACGAAA
CAAAGACGTAGAGTTGAGCAAGCAGGGTCAGGCAAAGCGTGGAGAGCCGGCTGAGTCTAGGTAGGCTC
CAAGGGAGCGCCGGACAAAGGCCCGGTCTCGACCTGAGCTTTAAACTTACCTAGACGGCGGACGCAGT
TCAGGAGGCACCACAGGCGGGAGGCGGCAGAACGCGACTCAACCGGCGTGGATGGCGGCCTCAGGTAG
GGCGGCGGGCGCGTGAAGGAGAGATGCGAGCCCCTCGAAGCTTCAGCTGTGTTCTGGCGGCAAACCCG
TTGCGAAAAAGAACGTTCACGGCGACTACTGCACTTATATACGGTTCTCCCCCACCCTCGGGAAAAAG
GCGGAGCCAGTACACGACATCACTTTCCCAGTTTACCCCGCGCCACCTTCTCTAGGCACCGGTTCAAT
TGCCGACCCCTCCCCCCAACTTCTCGGGGACTGTGGGCGATGTGCGCTCTGCCCACTGACGTCGAC-
3'
```

6.3.2 Peptide sequences of LDCV constructs

The key for all following amino acid colors is as follows:

- Green: hydrophobic uncharged residues
- Red: acidic residues
- Blue: basic residues
- Black: other residues

Sequences translated and analyzed by Peptide2.0

Translated sequence of hPOMC sorting signal

The translated sequence of the hPOMC sorting signal used in this project. The positively charged residue (R) and the hydrophobic core (green) are marked in bold.

MPRSCCS**R**S**GALLLALLL**QASMEVRG

Translated sequence of CgA sorting signal

The translated sequence of the CgA sorting signal used in this project. The positively charged residue (R) and the hydrophobic core (green) are marked in bold.

```
MRSAAVLALLLCAGQVTALPVNSPMNKGDTEVMKCIVEVISDTLSKPSPMPVSQECFET
LRGDERILSILRHQNLLKELQDLALQGAKERAHQQKKHSGFEDELSEVLENQSSQAELK
EAVEEPSSKDVMEKREDSKEAEKSGEATDGARPQALPEPMQESKAEGNNQAPGEEEEEE
EEATNTHPPASLPSQKYPGPQAEGDSEGLSQGLVDREKGLSAEPGWQAKREEEEEEEEE
AEAGEEAVPEEGPTVVLNPHPSLGYKEIRKGESRSEALAVDGAGKPGAEEAQDPEGKG
EQEHSQQKEEEEEMAVVPQGLFRGGKSGELEQEEERLSKEWEDSKRWSKMDQLAKELTA
EKRLEGQEEEEDNRDSSMKLSFRARAYGFRGPGPQLRRGWRPSSREDSLEAGLPLQVRG
YPEEKKEEEGSANRRPEDQELESLSAIEAELEKVAHQLQALRRG
```

Translated sequence of SgII sorting signal

The translated sequence of the SgII sorting signal used in this project. The positively charged residue (K) and the hydrophobic core (green) are marked in bold.

```
MAEAKTHWLGAALSLIPLIFLISGAEAASFQRNQLLQKEPDLRLENVQKFPSPEMIRALEYI
ENLRQQAHKEESSPDYNPYQGVSVPLQQKENGDESHLPERDSLSEEDWMRIILEALRQAENE
PQSAPKENKPYALNSEKNFPMDMSDDYETQQWPERKLKHMQFPPMYEENSRDNPFKRTNEIV
EEQYTPQSLATLESVFQELGKLTGPNNQKRERMDEEQKLYTDDEDDIYKANNIAYEDVVGGE
DWNPVEEKIESQTQEEVRDSKENIGKNEQINDEMKRSGQLGIQEEDLRKESKDQLSDDVSKV
IAYLKRLVNAAGSGRLQNGQNGERATRLFEKPLDSQSIYQLIEISRNLQIPPEDLIEMLKTG
EKPNGSVEPERELDLPVDLDDISEADLDHPDLFQNRMLSKSGYPKTPGRAGTEALPDGLSVE
DILNLLGMESAANQKTSYFPNPYNQEKVLPRLPYGAGRSRSNQLPKAAWIPHVENRQMAYEN
LNDKDQELGEYLARMLVKYPEIINSNQVKRVPGQGSSEDDLQEEEQIEQAIKEHLNQGSSQE
TDKLAPVSKRFPVGPPKNDDTPNRQYWDEDLLMKVLEYLNQEKAEKGREHIAKRAMENM
```

6.4 Sanger sequencing alignments

The following section contains the Sanger sequencing alignments after each cloning step of the eCas9-sgRNA plasmid and the pAAVS1-P-MCS donor plasmid. Furthermore, this section contains the Sanger sequencing alignments for the homology arms and donor DNA for the IMR90-4 hPOMC-GLuc clone 6, IMR90-4 hPOMC-GLuc clone 4, and IMR90-4 VAMP2-GLuc clone 11. In the alignments, the top row is the expected reference sequence and the remaining rows are the output of Sanger sequences of each plasmid or gDNA segment. The specific details for each sequence are described in each segment. Any bases marked in red were mismatched to the reference sequence.

6.4.1 sgRNA insertion confirmation

Sanger sequencing alignment for eCas9-sgRNA confirmation. 43EH20: Plasmid 2, 43EH19: Plasmid 1, 43EH21: Plasmid 3 (1 page).

Figure 112: Sanger sequencing alignments for eCas9-sgRNA plasmid. Any mismatched bases are indicated in red (1 page).

6.4.2 Donor plasmid with Ef1α promoter

Sanger sequencing alignment for Ef1a promoter. 43EH15: Plasmid 1, forward sequencing run; 43EH14: Plasmid 1, reverse sequencing run (3 pages).

```
                 1                                                                                              82
efla_953         GCCACTCCCACTGTCCTTTCCTPAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTCATCATTCTTATTCTGGGGGGTG
43EH15  (4....    ------------------------------------------------------------------------------------
43EH4   (4....    ------------------------------------------------------------------------------------

                 83                                                                                             164
efla_953         GGGTGGGGCAGGACAGCAAGGGGGAGAGGATTGGGGAAGACAATAGCAGGCGATAGCAGGCAGTAGACGTGGGGTCTTATGGACTAGTA
43EH15  (4....    ------------------------------------------------------------------------------------
43EH4   (4....    ------------------------------AGACAATAGCAGGCATGCTGGGATGGGGTCTTATGGACTAGTA

                 165                                                                                            246
efla_953         TTAATTAATTCAGCACCTGAAAATGAAGAAGAAGAAAAACTTTGAACCACACTGTCTGAGGCTTGAGAATGAACCACAAGATCCAAAC
43EH15  (4....    TTAATTAATTCAGCACACTGAAAATGGAAGAAAAAAACTTTGAACCACACTGTCTGAGGCTTGAGAATGAACCAAGATCCAAAC
43EH4   (4....    ----------------------------------------------------.........

                 247                                                                                            328
efla_953         TCAAAAAGGGCAAATTCCAAGAGGAGAATTACATCAAGTGCCAAGCTGGGCCTTACTTCAGTCTCCACCACTCAGTGTGGGGAA
43EH15  (4....    TCAAAAAGGGCAAATTCCAAGAGGAGAATTACATCAAGTGCCAAGCTGGGCCTTACTTCAGTCTCCACCACTCAGTGTGGGGAA
43EH4   (4....    GAACTCTCCCACCCACTTCCAACCCGAAAGCTCGGGGATCAAGAATCACGTACTTCCAGCCAGGGGCGTGGAAGTAATTCAAGGC

                 329                                                                                            410
efla_953         ACTCCATCGCATAAAACCCCTCCCCCCCCCAACCTAAACGACCTACTCCAAAAAGCTCGAGAACTAATCGAGGTGCCTGGACGG
43EH15  (4....    ACTCCATCGCATAAAACCCCTCCCCCCCCCAACCTAAACGACGACTACTCCAAAAAGCTCGAGAACTAATCGAGGTGCCTGGACGG
43EH4   (4....    ----------------------------------------------------.........

                 411                                                                                            492
efla_953         CGCCCGGTTACTCCGTGGAGTCACATGAAGCGACGGGCTGAGGACGGAAAAGGCCCTTTTCCTTTGTGTGTGGGGTGACTCACCCGCC
43EH15  (4....    CGCCCGGTTACTCCGTGGAGTCACATGAAGCGACGGGCTGAGGACGGAAAAGGCCCTTTTCCTTTGTGTGTGGGGTGACTCACCCGCC
43EH4   (4....    ----------------------------------------------------.........

                 493                                                                                            574
efla_953         CGCTCTCCCGAGCGCGCGGTCCTCTCCATTTTGAGCTCCCTGCACGCAGGGCGCGGGAAGCGGCCCATCTTTCCGCTCACGCAACTG
43EH15  (4....    CGCTCTCCCGAGCGCGCGGTCCTCTCCATTTTGAGCTCCCTGCAGCAGGGCGCGGGAAGCGGCCATCTTTCCGCTCACGCAACTG
43EH4   (4....    ----------------------------------------------------.........

                 575                                                                                            656
efla_953         GTGCCGACCGGGCCAGCCTTGCGCGCCAGGGGCGGTACACGCGCGCGCGAGGCGCCAGGCACCAGAGCAGGGCCGGCCAGC
43EH15  (4....    GTGCCGACCGGGCCAGCCTTGCGCGCCAGGGGCGGTACACGCGCGGCGCGAGGCGCCAGGCACCAGAGCAGGGCCGGCCAGC
43EH4   (4....    ----------------------------------------------------.........

                 657                                                                                            738
efla_953         TTGAGACTACCCCCGTCCGATTCTCGGGTGGGCGCGGCCGGGCGTCCGCAGGCCCCGGCCCCGCTCGCCAGGCCCCGCCGCCATCGGGCC
43EH15  (4....    GTGAGACTACCCCGTCCGATTCTCGGGTGGGCGCGGCCGGGCGTCCGCAGGCCCCGGCCCCGCTCGCCAGGCCCCGCCGCCATCGGGCC
43EH4   (4....    TTGAGACTACCCCCGTTCCGATTCTCGGGTTGGCACGGGTTGGGGACGCGCAGGGATTCCCCAA
```

```
                 739                                                                                            820
efla_953         CCGTCGCGGCCCCGGCCCCGGCCCCAAAAACCGAAATACCAGTGTGTCGAGATCTTGGCCCGGCATTTACACGACTATCTTGCCAGAAAA
43EH15  (4....    CCGTCGCGGCCCCGGCCCCGGCCCCAAAAACCGAAATACCAGTGTGTCGAGATCTTGGCCCGGCATTTACAAGACTATCTTGCCAGAAAA
43EH4   (4....    ----------------------------------------------------.........

                 821                                                                                            902
efla_953         AAAGCGTCGCAGCAGGTCATCAAAAAATTTTAAATGCTAGAGACTTATCGAAAAGACGAGCGAGAGACAGCGGCGAAAGGTGCCACCA
43EH15  (4....    AAAGCGTCGCAGCAGGTCATCAAAAAATTTTAAATGCTAGAGACTTATCGAAAAGACGAGCGAGAGACAGCGGCGAAAGGTGCCACCA
43EH4   (4....    ----------------------------------------------------.........

                 903                                                                                            984
efla_953         GATTCGCACGCGGCGCCCCAGCGGCCCGCCCAAGCCGAGGCCTCAACTCAGACGCGAGGCGAGGAAAGGGGCTCTTAAGCGCAAGGCCTC
43EH15  (4....    GATTCGCACGCGGCGCCCCAGCGGCCCGCCCAAGCCGAGGCCTCAACTCAGACGCGAGGCGAGGAAAGGGGCTCTTAAGCGCAAGGCCTC
43EH4   (4....    ----------------------------------------------------.........

                 985                                                                                            1066
efla_953         GAACTCTCCCACCCACTTCCAACCCGAAAGCTCGGGGATCAAGAATCACGTACTTCCAGCCAGGGGCGTGGAAGTAATTCAAGGC
43EH15  (4....    GAACTCTCCCACCCACTTCCAACCCGAAAGCTCGGGGATCAAGAATCACGTACTTCCAGCCAGGGGCGTGGAAGTAATTCAAGGC
43EH4   (4....    ----------------------------------------------------.........

                 1067                                                                                           1148
efla_953         ACGCAAGGGCCCCATAACCCGTAAAGAGAGGGCCCCGGGCGCGCGCGGAACCACACGCACGGCGCACTTACTGTGTTCTGGCGCGAAACCCGT
43EH15  (4....    ACGCAAGGGCCCCATAACCCGTAAAGAGAGGGCCCCGGGCGCGCGCGGAACCACACGCACGGCGCACTTACTGTGTTCTGGCGCGAAACCCGT
43EH4   (4....    --------------GTAAAGAGGGCCAGGCCCCGCGGGAACCACACGCGCACTTACTGTGTTCTGGCGCGCAAACCCGT

                 1149                                                                                           1230
efla_953         TGCGAAAAAGAAACGTTCACGGCGACTACTGCACTTATATACGTTCTCCCCCACCCTGGGGAGAAAAGGGGGGGAGCCGCCATGACTACAC
43EH15  (4....    TGCGAAAAAGAAACGTTCACGGCGACTACTGCACTTATATACGTTCTCCCCCACCCTGGGGAGAAAAGGGGGGGAGCCGCCATGACTACAC
43EH4   (4....    TGCGAAAAAGAAACGTTCACGGCGACTACTGCACTTATATACGTTCTCCCCCACCCTGGGGAGAAAAGGGGGGGAGCCAGTACAC

                 1231                                                                                           1312
efla_953         GACATCACTTTCCCAGTTTTACCCCCGCGCCACCTTCTTCTAGGCACCGGTTCAATTGCCGACCCCGTTCCCCCCAACTTCTTCCGGGG
43EH15  (4....    ---------------------------------------------------------GACCCCGTTCCCCCCAACTTCTTCCGGGG
43EH4   (4....    GACATCACTTTCCCAGTTTTACCCCCGCGCCACCTTCTTCTAGGCACCGGTTCAATTGCCGACCCCGTTCCCCCCAACTTCTTCTCGGGG

                 1313                                                                                           1394
efla_953         ACTGTGGCGCGATGTGCGCTCTCGCCACTGACGCGGACGCACCGGAGACCTTCACGGCATACTCTTCTCCCAACTTCAGTGATGACGGAGAGC
43EH15  (4....    ----------------------------------------------------.........
43EH4   (4....    ACTGTGGCGCGATGTGCGCTCTCGCCACTGACGCGGACGCACCGGAGACCTCACGGCATACTCTTCTCCCAACTTCAGTGATGACGGAGAGC

                 1395                                                                                           1476
efla_953         GGGCGGGGTGAGGGGCGGGGGAACGCAGCCGATCTCTGGGTTCTACGTTAGTGGGGAGTTTAACGACGGGTCCCTGGGGATTCCCCAA
43EH15  (4....    GGGCGGGGTGAGGGGGCGGGGGAACGCAGCGATCTCTGGGTTCTACGTTAGTGGGAGTTTAACGACGGGTCCCTGGGGATTCCCCAA
43EH4   (4....    GGGCGGGGTGAGGGGCGGGGGAACGCAGCCGATCTCTGGGTTCTACGTTAGTGGGGAGTTTAACGACGGGTCCCTGGGGATTCCCCAA
```

213

Section 6.4

```
Page 3  43EH15 (43EH15.abl), 43EH14 (43EH14.abl)

                1477                                                                                      1558
efla_953        GGCAGGGGCGGAGTCCTTTTGTTATGAATTACTCTCAGCTCCGGTCCGGCGGCGGCGGGTTGGGGGGGGGTTGGTTGACGGGGAGGCCGGCCT
43EH15 (4...    --------------------------------------------------------------------------------------
43EH14 (4...    GGCAGGGGCCAGTCCTTTTTGTTATGAATTACTCTCAGCTCCGGTCCGGGGCGGGTGGGGGGGGGGTTGGGGGGAGGCCGGCCT
                ........................................................................................

                1559                                                                                      1640
efla_953        GGAAGGGACGTGCAGAATCTTCCCTCTACCATTGCTGGCTTAGCTCCAAAGGTTGTATTGAGATTAGGGTGTACCTTCGCCT
43EH15 (4...    --------------------------------------------------------------------------------------
43EH14 (4...    GGAAGGGACGTGCAGAATCTTCCCTCTACCATTGCTGGCTTAGCTCCAAAGGTTGTATTGAGATTAGGGTGTACCTTCGCCT
                ........................................................................................

                1641                                                                                      1722
efla_953        CTCAATCAGCCTCCCGTCCTCAGCCTTGCCATCTCGTGAGTCCGGGACAAATCCCTAGAGCGTCTTCCTCTGCCGGGGTCTCAG
43EH15 (4...    --------------------------------------------------------------------------------------
43EH14 (4...    CTCAATCAGCCTCCCGTCCTCAGCCTTGCCATCTCGTGAGTCCGGGACAAATCCCTAGAGCGTCTTCCTCTGCCGGGGTCTCAG
                ........................................................................................

                1723                                                                                      1804
efla_953        CCCAGCCCCGGGGTCGACAGTACTAAGCTTTGACAGAAAAGCCCCATCCTTAGGCCTCCTCCTTCCTAGTCTCCTGATATTGG
43EH15 (4...    --------------------------------------------------------------------------------------
43EH14 (4...    CCCAGCCCCGGGGTCGACAGTACTAAGCTTTGACAGAAAAGCCCCATCCTTAGGCCTCCTCCTTCCTAGTCTCCTGATATTGG
                ........................................................................................

                1805                                                                                      1886
efla_953        GTCTAACCCCACCTCCTGTTAGGCAGATTCCTTATCTGGTGACACACCCCATTTCCTGGAGCCATCTCTCTCCTTGCCAG
43EH15 (4...    --------------------------------------------------------------------------------------
43EH14 (4...    GTCTAACCCCACCTC----------------------------------------------------------------------
                ........................................................................................

                1887                                                                                      1968
efla_953        AACCTCTAAGGTTTGCTTACGATGGAGCCAGAGAGGATCCTGGGCAGCGGGGGTGGGAGGGAGGAGGGGGGGGA
43EH15 (4...    --------------------------------------------------------------------------------------
43EH14 (4...    --------------------------------------------------------------------------------------
                ........................................................................................

                1969
efla_953        TGCG
43EH15 (4...    ---
43EH14 (4...    ---
```

Figure 113: Sanger sequencing alignments for donor plasmid with Ef1α promoter. Any mismatched bases are indicated in red (3 pages).

6.4.3 Donor plasmid with Ef1-HTLV promoter

Sanger sequencing alignment for Ef1-HTLV promoter. 43EH18: Plasmid 5, 43EH17: Plasmid 6 (1 page).

Figure 114: Sanger sequencing alignments for donor plasmid with Ef1-HTLV promoter. Any mismatched bases are indicated in red (1 page).

6.4.4 Donor plasmid with Ef1α promoter and hPOMC-GLuc

Sanger sequencing alignment for Ef1a promoter plus hPOMC-GLuc. 43EH24: Plasmid 6, 43EH22: Plasmid 3, 43EH23: Plasmid 4 (2 pages).

Page 1 43EH24 (43EH24.abl), 43EH22 (43EH22.abl), 43EH23 (43EH23.abl)

```
                79                                                                                        160
959       ctagtcaacaccggccccccttgatcttgtccacctgccctggaatctgctggcaaagtcggacaagcggttcgggcagcac
43EH24 (4....  CTAGTCACACCGGCCCCCCTTGATCTTGTCCACCTGCCCTGGAATCTTGCTGGCAAAGGTCGCACAGCGTTCGCGGCAGCCAC
43EH22 (4....  CTAGTCACACCGGCCCCCCTTGATCTTGTCCACCTGCCCTGGAATCTTGCTGGCAAAGGTCGCACAGCGTTCGCGGCAGCCAC
43EH23 (4....  CTAGTCACACCGGCCCCCCTTGATCTTGTCCACCTGCCCTGGAATCTTGCTGGCAAAGGTCGCACAGCGTTCGCGGCAGCCAC
..........

                161                                                                                       242
959       ttcttgagcaaggtcaagaacactgcacggttgggcaagcccttgaggcaggccagtgtgcaagtccacacaagatcgaacctgtg
43EH24 (4....  TTCTTGAGCAGGTCAGAACACTGCACGTTGGGCAAGCCCTTTGAGGCAGGCCAGTGTGCAGTCCACACACAAGATCGAACCTGTG
43EH22 (4....  TTCTTGAGCAGGTCAGAACACTGCACGTTGGGCAAGCCCTTTGAGGCAGGCCAGTGTGCAGTCCACACACAAGATCGAACCTGTG
43EH23 (4....  TTCTTGAGCAGGTCAGAACACTGCACGTTGGGCAAGCCCTTTGAGGCAGGCCAGTGTGCAGTCCACACACAAGATCGAACCTGTG
..........

                243                                                                                       324
959       cgatgaaactgctccaagtcctgaaccccaggaatctcaggaatgtcgacatcgcctcgctcatgccgccctg
43EH24 (4....  CGATGAACTGCTCCATGGGGCTCCCAAGTCCTTGAACCCAGGAATCTCAGGAATGTCGACATCGCCTCGCCTATGCCGCCCCTG
43EH22 (4....  CGATGAACTGCTCCATGGGGCTCCCAAGTCCTTGAACCCAGGAATCTCAGGAATGTCGACATCGCCTCGCCTATGCCGCCCCTG
43EH23 (4....  CGATGAACTGCTCCATGGGGCTCCCAAGTCCTTGAACCCAGGAATCTCAGGAATGTCGACATCGCCTCGCCTATGCCGCCCCTG
..........

                325                                                                                       406
959       tgccagaactcctttgtcgccttcgtaggtgtggcaagcgtcctgggatgaacttcttcatcttgggcgtgaacttgatgtgaggac
43EH24 (4....  TGCGGACTCTTTGTCGCCCTTCGTAGGTGTGGCAGCGTCCTGGGATGAACTTCTTCATCTTGGGCGTGCACTTGATGTGTGGAC
43EH22 (4....  TGCGGACTCTTTGTCGCCCTTCGTAGGTGTGGCAGCGTCCTGGGATGAACTTCTTCATCTTGGGCGTGCACTTGATGTGTGGAC
43EH23 (4....  TGCGGACTCTTTGTCGCCCTTCGTAGGTGTGGCAGCGTCCTGGGATGAACTTCTTCATCTTGGGCGTGCACTTGATGTGTGGAC
..........

                407                                                                                       488
959       aaggcagatcagacaacagccccctggtcgcagccagcttccgggcaatccgtgcgcaaggccacagctgttgaagcgcagtcct
43EH24 (4....  AGGCAGATCAGACAGCCCCCTGGTCGCAGCCAGCTTTCCGGGCAATTGGCTTCCATCTCTTTGAGCACACTTCCAGCGCGCAGCTTCT
43EH22 (4....  AGGCAGATCAGACAGCCCCCTGGTCGCAGCCAGCTTTCCGGGCAATTGGCTTCCATCTCTTTGAGCACACTTCCAGCGCGCAGCTTCT
43EH23 (4....  AGGCAGATCAGACAGCCCCCTGGTCGCAGCCAGCTTTCCGGGCAATTGGCTTCCATCTCTTTGAGCACACTTCCAGCGCGCAGCTTCT
..........

                489                                                                                       570
959       tgccgggcaacttcccgcgggtcagcaatccgatggtgcgcaaggccagacagatgttgaagtcttcgtt
43EH24 (4....  TGCCGGGCAACTTCCCGCGGGTCAGCATCGGATCCGTTGGTCGTCGCGCAAGGTGGTCGGCCAGACGGATGTTGAAGTCTTCGTT
43EH22 (4....  TGCCGGGCAACTTCCCGCGGGTCAGCATCGGATCCGTTGGTCGTCGCGCAAGGTGGTCGGCCAGACGGATGTTGAAGTCTTCGTT
43EH23 (4....  TGCCGGGCAACTTCCCGCGGGTCAGCATCGGATCCGTTGGTCGTCGCGCAAGGTGGTCGGCCAGACGGATGTTGAAGTCTTCGTT
..........

                571                                                                                       652
959       gtttcctcggtgggcttcatgccacgcaacttccatggaggcctgaagcgccaacaagggcccccgagcggcggctgcag
43EH24 (4....  GTTCTCGGTGGGCTTCATGCCACGCACACCACCACCTGAAATGGAGAGAAAAAAAACTTTGAACCACTGTCTGAGGCTTGAG
43EH22 (4....  GTTCTCGGTGGGCTTCATGCCACGCACACCACCACCTGAAATGGAGAGAAAAAAAACTTTGAACCACTGTCTGAGGCTTGAG
43EH23 (4....  GTTCTCGGTGGGCTTCATGCCACGCACACCACCACCTGAAATGGAGAGAAAAAAAACTTTGAACCACTGTCTGAGGCTTGAG
..........

                653                                                                                       734
959       cacgatccggcatggtggTTAATTATTCACGACACTGGAATGGAGAGAAAAAAAACTTTGAACCACTGTCTGAGGCTTGAG
43EH24 (4....  CACGATCTCGGCATGGTGGTTAATTATTCACGACACCTGAAATGGAGAGAAAAAAAACTTTGAACCACTGTCTGAGGCTTGAG
43EH22 (4....  CACGATCTCGGCATGGTGGTTAATTATTCACGACACCTGAAATGGAGAGAAAAAAAACTTTGAACCACTGTCTGAGGCTTGAG
43EH23 (4....  CACGATCTCGGCATGGTGGTTAATTATTCACGACACCTGAAATGGAGAGAAAAAAAACTTTGAACCACTGTCTGAGGCTTGAG
..........
```

Page 2 43EH24 (43EH24.abl), 43EH22 (43EH22.abl), 43EH23 (43EH23.abl)

```
                735                                                                                       816
959       AATGAACCAAGATCCAAACTCAAAAAGGGAGCCAAATTCCAAGGAGAATTACATCAGGTGCCAAGCTGGCCTAACTTCAGTCTCC
43EH24 (4....  AATGAACCAAGATCCAAACTCAAAAAGGGAGCCAAATTCCAAGGAGAATTACATCAGGTGCCAAGCTGGCCTAACTTCAGTCTCC
43EH22 (4....  AATGAACCAAGATCCAAACTCAAAAAGGGAGCCAAATTCCAAGGAGAATTACATCAGGTGCCAAGCTGGCCTAACTTCAGTCTCC
43EH23 (4....  AATGAACCAAGATCCAAACTCAAAAAGGGGCCAAATTCCAAGAGAGAATTACATCAAGGTGCCAAGCTGGCCTAACTTCAGTCTCC
..........

                817                                                                                       898
959       ACCCACTCAGTGTGGGAGAAACTCCATGCATAAAAACCCCTTCCCCCCCAACCTAAAGACAAACGTACTCCAAAAGCTCGAGAACT
43EH24 (4....  ACCCACTCAGTGTGGGAGAAACTCCATGCATAAAAACCCCTTCCCCCCCAACCTAAAGACAAACGTACTCCAAAAGCTCGAGAACT
43EH22 (4....  ACCCACTCAGTGTGGGAGAAACTCCATGCATAAAAACCCCTTCCCCCCCAACCTAAAGACAAACGTACTCCAAAAGCTCGAGAACT
43EH23 (4....  ACCCACTCAGTGTGGGAGAAACTCCATGCATAAAAACCCCTCCCCCCCAACCTAAAGACAAACGTACTCCAAAAGCTCGAGAAACT
..........

                899                                                                                       980
959       AATCGAGGTGCCTGGACGGCGCCGGTACTCCGTGGAGTCACATGAAGCAGCAGCCTGAAGGAGCCGGAAAGCCCTTTTTCCTTTG
43EH24 (4....  AATCGAGGTGCCTGGACGGCGCCGGTACTCCGTGGAGTCACATGAAGCAGCAGCCTGAAGGAGCCGGAAAGCCCTTTTTCCTTTG
43EH22 (4....  AATCGAGGTGCCTGGACGGCGCCGGTACTCCGTGGAGTCACATGAAGCAGCAGCCTGAAGGAGCCGGAAAGCCCTTTTTCCTTTG
43EH23 (4....  AATCGAGGTGCCTGGACGGCGCCGGTACTCCGTGGAGTCACATGAAGCAGCGGCCTGAAGGAGCCGGAAAGCCCTTTTTCCTTTG
..........

                981                                                                                       1062
959       TGTGGGTGACTCACCCGCGCCGCTCTCCGAGCGGCCGGCGTCTCCAGTTTGAGCTCCCTGCAGCGAGGCCGGACCGGAAGCGGCCAT
43EH24 (4....  TGTGGGTGACTCACCCGCGCCGCTCTCCGAGCGGCCGGCGTCTCCAGTTTGAGCTCCCTGCAGCAGAGGCCGGGCGGAAGCGGCCAT
43EH22 (4....  TGTGGGTGACTCACCCGCGCCGCTCTCCGAGCGGCCGGCGTCTCCAGTTTGAGCTCCCTGCAGCAGAGGCCGGGCGGAAGCGGCCAT
43EH23 (4....  TGTGGGTGACTCACCCGCGCCGCTCTCCGAGCGGCCGGCGTCTCCAGTTTGAGCTCCCTGCAGCAGAGGCCGGGCGGAAGCGGCCAT
..........

                1063                                                                                      1144
959       CTTTCCGCTCAGCGACTACCCCGCCCCGCCCAGGCGCGGCGCCCAGCCTTGCCCGCCCAGGGCGGGGCGGATACACGGCGGCAGGCA
43EH24 (4....  CTTTCCGCTCAGCGACTACGTCGCCCCGCCCAGGCGCGGCGCCCAGCCTTGCCCGCCCAGGGCGGGGCGGATACACGGCGGCAGGCA
43EH22 (4....  CTTTCCGCTCAGCGACTACGTCGCCCCGCCCAGGCGCGGCGCCCAGCCTTGCCCGCCCAGGGCGGGGCGGATACACGGCGGCAGGCA
43EH23 (4....  CTTTCCGCTCAGCGACTACGTCGCCCCGCCCAGGCGCGGCGCCCAGCCTTGCCCGCCCAGGGCGGGGCGGATACACGGCGGCAGGCA
..........

                1145                                                                                      1195
959       CCAGAGCAGGCGCGGCCAGCTTGAGACTACCCCCCGTTCGGATTCTCGGTGGCC
43EH24 (4....  CCAGAGCAGGCGCGGCCAGCTTGAGACTACCCCCCGTTCGGATTCTCGGTGGCC
43EH22 (4....  CCAGAGCAGGCGCGGCCAGCTTGAGACTACCCCCCGTTCGGATTCTCGGTGGCC
43EH23 (4....  CCAGAGCAGGCGCGGCCAGCTTGAGACTACCCCCCGTTCGGATTCTCGGTGGCC
..........
```

216   Section 6.4

Figure 115: Sanger sequencing alignments for donor plasmid with Ef1α promoter and hPOMC-GLuc. Any mismatched bases are indicated in red (2 pages).

6.4.5 Donor plasmid with Ef1-HTLV promoter and hPOMC-GLuc

Sanger sequencing alignment for hPOMC-GLuc in Ef1-HTLV promoter plasmid. 43EH25: Plasmid 1, 43EH26: Plasmid 2, 43EH27: Plasmid 3 (1 page).

Figure 116: Sanger sequencing alignments for donor plasmid with Ef1-HTLV promoter and hPOMC-GLuc. Any mismatched bases are indicated in red (1 page).

6.4.6 Donor plasmid with Ef1-HTLV promoter and no tag GLuc

Sanger sequencing alignment for no tag GLuc in Ef1-HTLV promoter plasmid. 52BH06: Plasmid 1, 52BH05: Plasmid 3, 52BH04: Plasmid 7 (1 page).

Figure 117: Sanger sequencing alignments for donor plasmid with Ef1-HTLV promoter and no tag GLuc. Any mismatched bases are indicated in red (1 page).

6.4.7 Donor plasmid with Ef1-HTLV promoter and CgA-GLuc

Sanger sequencing alignment for CgA-GLuc in Ef1-HTLV promoter plasmid. 52BH88: Plasmid 1, forward sequencing run; EF0074784: Plasmid 7, forward sequencing run; EF0074785: Plasmid 3, forward sequencing run; EF0074784_2: Plasmid 1, middle sequencing run; EF0074785_2: Plasmid 3, middle sequencing run; EF0074785_3: Plasmid 1, reverse sequencing run; EF0074784_2: Plasmid 3, reverse sequencing run (5 pages).

```
                    1217                                                                                          1298
CgA-Gluc_...        ACTAGTctagtcaccacacggccccctggccccctgatcttgtccacctggccctggcgaaaggtcgcacaggttgcggc
52BH88_47...        ACTAGTCTAGTCACCACACCGGCCCCCTTGAGTCTTGTCCACCTGGCCCTGGATCGTTGTCGGCAAAGGTCGGCAAGGGTTGCGGC
EF0074784...        ACTAGTCTAGTCACCACCACCGGCCCGGTCCCCCTTGATCTTTGTCC●CCTGGCCCTGGATCGTTGTCGTGGCAAAGGTCGCACAGGGTTGCGGC
EF0074785...        ACTAGTCTAGTCACCACCACCGGCCCCCTTGATCTTTGTCC-------CTGGATCGTTGTCGTGGCAAAGGTCGCACAGGGTTGCGGC
EF0074784...        --------------------------------------------------------------
EF0074785...        --------------------------------------------------------------
EF0074784...        --------------------------------------------------------------
EF0074785...        --------------------------------------------------------------
EF0074784...        --------------------------------------------------------------
EF0074785...        --------------------------------------------------------------
EF0074784...        --------------------------------------------------------------
.................................................................................

                    1299                                                                                          1380
CgA-Gluc_...        agccacttcttgagcagtcaagaacactgcacgttggcaagccctctgaggcagccaagtgtgagtccacacacagatcga
52BH88_47...        AGCCACTTCTTGAGCAGGTCAGAACACTGCACGTTGGCCAAGCCCTTTGAGGCAGCCAGTTGTGCAGTCCACACACAGATCGA
EF0074784...        AGCCACTTCTTGAGCAGGTCAGAACACTGCACGTTGGCCAAGCCCTTTGAGGCAGCCAGTTGTGCAGTCCACACACAGATCGA
EF0074785...        AGCCACTTCTTGAGCAGGTCAGAACACTGCACGTTGGCCAAGCCCTTTGAGGCAGCCAGTTGTGCAGTCCACACACAGATCGA
EF0074784...        --------------------------------------------------------------
EF0074785...        --------------------------------------------------------------
EF0074784...        --------------------------------------------------------------
EF0074785...        --------------------------------------------------------------
EF0074784...        --------------------------------------------------------------
EF0074785...        --------------------------------------------------------------
EF0074784...        --------------------------------------------------------------
.................................................................................

                    1381                                                                                          1462
CgA-Gluc_...        cctgtgcgatgaactgctccatggctccaagtcttgaacccaggaatctcaggaatgtcgaacgatcgcctcgcctatgcc
52BH88_47...        CCTGTGCGATGAACTGCTCCATGGCTCCAAGTCTTGAACCCAGGAATCTCAGGAATGTCGAACGATCGCCTCGCCTATGCC
EF0074784...        CCTGTGCGATGAACTGCTCCATGGGCTCCAAGTCTTGAACCCAGGAATCTCAGGAATGTCGAACGATCGCCTCGCCTATGCC
EF0074785...        CCTGTGCGATGAACTGCTCCATGGGCTCCAAGTCTTGAACCCAGGAATCTCAGGAATGTCGAACGATCGCCTCGCCTATGCC
EF0074784...        --------------------------------------------------------------
EF0074785...        --------------------------------------------------------------
EF0074784...        --------------------------------------------------------------
EF0074785...        --------------------------------------------------------------
EF0074784...        --------------------------------------------------------------
EF0074785...        --------------------------------------------------------------
EF0074784...        --------------------------------------------------------------
.................................................................................

                    1463                                                                                          1544
CgA-Gluc_...        gccctgtgcggactccttgtcgccttcgtaggtgtggcaagcgtcctgggatgaacttcttcatcttgggcgtgcacttgatg
52BH88_47...        GCCCTGTGCGGACTCCTTTGTTGTCGCCTTCGTAGGTGTGGCAGCGTCCTGGGATGAACTTCTTCATCTTTGGGCGTGCACTTGATG
EF0074784...        GCCCTGTGCGGACTCCTTTGTCGCCTTCGTAGGTGTGGCAGCGTCCTGGGATGAACTTCTTCATCTTTGGGCGTGCACTTGATG
EF0074785...        GCCCTGTGCGGACTCTTTGTCGCCTTCGTAGGTGTGGCAGCGTCCTGGGATGAACTTCTTCATCTTTGGGCGTGCACTTGATG
EF0074784...        --------------------------------------------------------------
EF0074785...        --------------------------------------------------------------
EF0074784...        --------------------------------------------------------------
EF0074785...        --------------------------------------------------------------
EF0074784...        --------------------------------------------------------------
EF0074785...        --------------------------------------------------------------
EF0074784...        --------------------------------------------------------------
.................................................................................

                    1545                                                                                          1626
CgA-Gluc_...        tgggacaaggcaggagagatcaagacagacccctggtgcagcacagcttccgggggattggcttccatctatctttgagcaactccaggcgca
52BH88_47...        TGGGACAAGGCAGGAGAGATCAAGACAGACCCCTGGTGCAGCACAGCTTCCGGGGATTGGCTTCCATCTCTTTGAGCAACTCCCAGCGGCA
EF0074784...        TGGGACAAGGCAGGAGAGATCAAGACAGACCCCTGGTGCAGCACAGCTTCCGGGGCATTGGCTTCCATCTCTTTGAGCAACTCCCAGCGGCA
EF0074785...        TGGGACAAGGCAGGACTCAGACAGACCCCTGGTGCAGCACAGCTTCCGGGGCACTTCCGGGGCATTGGCTTCCATCTCTTTGAGCACCTCCCAGCGGCA
EF0074784...        ----|---|---------------------------------------
EF0074785...        ----|---|---------------------------------------
EF0074784...        ----|---|---------------------------------------
EF0074785...        ----|---|---------------------------------------
EF0074784...        ----|---|---------------------------------------
EF0074785...        ----|---|---------------------------------------
EF0074784...        ----|---|---------------------------------------
.................................................................................
```

```
                    1627                                                                                          1708
CgA-Gluc_...        gctcttgccgtggcaacttcccgcggtcagcatcagagatcccgtggtcgcgaaagttgctggccacggccacgatgttgaagtc
52BH88_47...        GCTTCTTGCCGGGGCAACTTCCCGCGGTCACATCAGAGATCC---------------
EF0074784...        GCTTCTTGCCGGGGCAACTTCCCGCGGTCACATCAGAGATCC---------------
EF0074785...        GCTTCTTGCCGGGGCAACTTCCCGCGGTCAACATCCAGAGATCCGTGGTCGCGGAAAGTTGCTGGCCACGGCCACGATGTTGAAGTC
EF0074784...        ---------------------------------------------------------
EF0074785...        ---------------------------------------------------------
EF0074784...        ---------------------------------------------------------
EF0074785...        ---------------------------------------------------------
EF0074784...        ---------------------------------------------------------
EF0074785...        ---------------------------------------------------------
EF0074784...        ---------------------------------------------------------
.................................................................................

                    1709                                                                                          1790
CgA-Gluc_...        ttcgttgttctcggtgggcttcattggatccCAGCTCGGTACCggccccgccgcctagtgcctgcgaagctggtgggccacttc
52BH88_47...        TTCGTTGTTCTCGGTGGGCTTCATTTGGATCCGAGCTCGGTACCGGCCCCGCCGCCTAGTGCCTGCGAAGCTGGTGGGCCACTTTC
EF0074784...        TTCGTTGTTCTCGGTGGGCTTCATTGGATCCGAGCTCGGTACCGGCCCCGCCGCCTAGTGCCTGCGAAGCTGGTGGGCCACTTTC
EF0074785...        --------------------------------------------------------
EF0074784...        ----CCCGCCGTAGTGCCTGCGAAGCTGGTGGGCCACTTTC
EF0074785...        ----CCCGCCGTAGTGCCTGCGAAGCTGGTGGGCCACTTTC
EF0074784...        ----CCCGCCGTAGTGCCTGCGAAGCTGGTGGGCCACTTTC
EF0074785...        ---------------------------------------------------------
EF0074784...        ---------------------------------------------------------
.................................................................................

                    1791                                                                                          1872
CgA-Gluc_...        tccagctctgctcaatggccgacaggctctccagctcctgtcctctggtctctgcggttcggctgcctcctcctcttt
52BH88_47...        ---------------------------------------------------------
EF0074784...        TCCAGCTCTGC●GCTCAATGGCCGACAGGCTCTCCAACTCCTGGTCCTCTGGTCTCTGCGGTTCGGCTGCCTCCTCCTCTTTCT
EF0074785...        TCCAGCTCTGC●GCTCAATGGCCGACAGGCTCTCCAACTCCTGGTCCTCTGGTCTCTGCGGTTCGGCTGCCTCCTCCTCTTTCT
EF0074784...        TCCAGCTCTGG●GTCAATGGCCGACAGGCTCTCCAACTCCTGGTCCTCTGGTCTCTGCGGTTCGGCTGCCTCCTCCTCTTTCT
EF0074785...        TCCAGCTCTGG●GTCAATGGCCGACAGGCTCTCCAACTCCTGGTCCTCTGGTCTCTGCGGTTCGGCTGCCTCCTCCTCTTTCT
EF0074784...        ---------------------------------------------------------
.................................................................................

                    1873                                                                                          1954
CgA-Gluc_...        tctcctcagggtagcctcggacctggaggggcaaggcccgcctcaagcgtgtcctcccggggaatggcctccaagcctcgtcg
52BH88_47...        ---------------------------------------------------------
EF0074784...        TCTCCTCGGGGTAGCCTCGGACCTGGAGGGGCAG-----------------------
EF0074785...        TCTCCTCGGGGTAGCCTCGGACCTGGAGGGGCAGGCCGCCGCCTCAAGCGTGTCCTCCCCGGGGAAGAGGCCTCCAAGCCTCGTCG
EF0074784...        TCTCCTCGGGGTAGCCTCGGACCTGGAGGGGCAGGCCGCCGCCTCAAGCGTGTCTCTCCCCGGGGAAGAGGCCTCCAAGCCTCGTCG
EF0074785...        ---------------------------------------------------------
EF0074784...        ---------------------------------------------------------
.................................................................................

                    1955                                                                                          2036
CgA-Gluc_...        cagtcgcggcccagggcccctgaagccgtaggccccggaaagagaagcttcatggaactgtcccggttgtcctcctcc
52BH88_47...        ---------------------------------------------------------
EF0074784...        CAGCTCGCGCCCAGGGCCCCTGAAACCGTAGGCCCCGGGCCCCTGAAGAGAAGCTTCATGGAACTGTCCCGGTTGTCTCCTCTCC
EF0074785...        CAGCTCGCGCCCAGGGCCCCTGAAACCGTAGGCCCCGGGCCCCTGAAGAGAAGCTTCATGGAACTGTCCCGGTTGTCTCCTCTCC
EF0074784...        CAGCTCGCGCCCAGGGCCCCTGAAGCCCTCAAGCCCCGGGCCCCTGGAAGGAGGAGCTTCATGGAACTGTCCCGGTTGTCTCCTCTCC
EF0074785...        ---------------------------------------------------------
EF0074784...        ---------------------------------------------------------
.................................................................................
```

Section 6.4

```
                2037                                                                                                          2118
CgA-GLuc_...    tcctctgtcccctccagccgcttctcagccgtcagctccttggccagctcggtccatcttgctccaggcgttggagtcctcc
52BH88_47...    ----------------------------------------------------------------------------------
EF0074784...    ----------------------------------------------------------------------------------
EF0074784...    ----------------------------------------------------------------------------------
EF0074785...    ----------------------------------------------------------------------------------
EF0074784...    TCCTCCTGCCCCTCCACCCGCTTCTCAGCCGTCAGCTCCTTGGCCAACTGGTCCATCTTGCTCCAGCGTTGGAGTCCTCC
EF0074785...    TCCTCCTGCCCCTCCACCCGGCTTCTCAGCCGTCAGCTCCTTGGGCCAGCTGGTCCATCTTGCTCCAGCGTTGGAGTCCTCCTCG
EF0074785...    --TTGCCTGCCACCCTGGGCTTGCACCTGGGCTCATCTTGCTCCAGCGTTTGGAGTCCTCC
EF0074784...    ----------------------------------------------------------------------------------
```

```
                2119                                                                                                          2200
CgA-GLuc_...    actccttggagaggccgctcctcctcctcgtccagctctccgttcttcccacacccgggaaggagggccttgcgggaccactgccat
52BH88_47...    --------------------------------------------------------------------------------------
EF0074784...    --------------------------------------------------------------------------------------
EF0074784...    --------------------------------------------------------------------------------------
EF0074785...    ----------------------------------------------------------------------------------------
EF0074784...    ACTCCTTGGAGAGGCCGGCTCCTCCTCCTCTTCCCGGCTCTTCCCACCCCGGAAGGAGGGCCTTGCGGGACCACTGCCAT
EF0074785...    ACTCCTTGGAGAGGCCGGCTCCTCCTCTTCCCGGCTCTTCCCACCCCGGAAGGAGGGCCTTGCGGGACCACTGCCAT
EF0074785...    ACTCCTTGGAGAGGCCGGCTCTCCTCCTCTTCCGGCTCTTCCCACCCCGGAAGGAGGGCCTTGCCGGGACCACTGCCAT
EF0074784...    --------------------------------------------------------------------------------------
```

```
                2201                                                                                                          2282
CgA-GLuc_...    ctctcctcctccttctgctggagtgtctgttctccttccctccctgggtcctgagcctcctcagcccaagcttc
52BH88_47...    ------------------------------------------------------------------------------
EF0074784...    ------------------------------------------------------------------------------
EF0074784...    ------------------------------------------------------------------------------
EF0074785...    ------------------------------------------------------------------------------
EF0074784...    CTCCTCCTCCTCCTTCTTCTGCTGGGGAGTGTCCTGTTCTCCCTCCCTTCCCTGGGGTCCTGAGCCTCCTCAGCCCAGGCTTC
EF0074785...    CTCTCCTCCTCCTCCTTTTTCTGCTGGGGAGTGCTCCGTTCTCCCTCCCTTCCCTGGGGTCCTGAGCCTCCTCAGCCCCAGGCTTC
EF0074785...    ------------------------------------------------------------------------------
EF0074784...    ------------------------------------------------------------------------------
```

```
                2283                                                                                                          2364
CgA-GLuc_...    ccagctccatccacagccaaggcctccgaccgactctcgcctttccggatctccttgtagccaaggctcgggtgggggttca
52BH88_47...    --------------------------------------------------------------------------------------
EF0074784...    --------------------------------------------------------------------------------------
EF0074784...    --------------------------------------------------------------------------------------
EF0074785...    --------------------------------------------------------------------------------------
EF0074784...    CCAGCTCCATCCACAGCCACAGCCTCCGACCACTCTCGCGCTTTCCGGATCTCCTTGTAGCCAAGGCTCGGGTGGGGGTTCA
EF0074785...    CCAGCTCCATCCACAGCCAGAGCCTTCCGGACCGACTCTCGCCTTTCCGGGATCTCCTTGTAGCCAAGGCTCGGGTGGGGGTTCA
EF0074785...    --------------------------------------------------------------------------------------
EF0074784...    --------------------------------------------------------------------------------------
```

```
                2365                                                                                                          2446
CgA-GLuc_...    gcactacagtggggccttctccttccctcggggacaagcctcctctccagcctcagcccctcctccctcctccctcctccctcttcct
52BH88_47...    ----------------------------------------------------------------------------------------------
EF0074784...    ----------------------------------------------------------------------------------------------
EF0074785...    ----------------------------------------------------------------------------------------------
EF0074784...    ----------------------------------------------------------------------------------------------
EF0074784...    GCACTACAGTGGGGCGCTTCTCCTTCCCTCGGGGACAGCCTCCTCTCCTCTCCAGCCTCAGCCTCCTCCTCCTCCTCCTCTCTCT
EF0074785...    GCACTACAGTGGGGGCCTTCTTCTCCTCGGGGACAGCCTCCTCTCTCCTCTCTCCAGCCTCCTCCTCCTCCTCCTCTCTCT
EF0074784...    GCACTACAGTGGGGCCTTCTTCCTCGGGGACAGCCTCTCACTTCTCCTCTCTCCTCCTCCTCCTCCTCTCTCT
EF0074784...    ----------------------------------------------------------------------------------------------
```

```
                2447                                                                                                          2528
CgA-GLuc_...    cttgactgccaacctggctcgcactcaaggccctctctcctgtccaccacagaccctgagagaggccctcaactgtccccctcg
52BH88_47...    ----------------------------------------------------------------------------------------
EF0074784...    ----------------------------------------------------------------------------------------
EF0074785...    ----------------------------------------------------------------------------------------
EF0074784...    CTTTGCCTGCCACCCTGGCTCTGCACTCAGGCCCCTTCTCTCTGTCCACCACAGACCCTGAGAGAGGCCCTCACTGTCCCCCTCG
EF0074785...    CTTTGCCTGCCACCCTGGCTCTGCACTCAGGCCCCTTCTCTCTGTCCACCACAGACCCTGAGAGAGGCCCTCACTGTCCCCCTCG
EF0074785...    CTTTGCCTGCCACCCTGGCTCTGCACTCAGGCCCCTTCTCTCTGTCCACCACAGACCCTGAGAGAGGCCCTCACTGTCCCCCTCG
EF0074784...    ----------------------------------------------------------------------------------------
```

```
                2529                                                                                                          2610
CgA-GLuc_...    gcctgtggcctgaggtatttctggctggggaggctggctggaggaggtggggtgttgtggtggctccctcctcctcctcctcctcct
52BH88_47...    ----------------------------------------------------------------------------------------------
EF0074784...    ----------------------------------------------------------------------------------------------
EF0074785...    GCCTGTGGGCCTGGGGTATTTCTGGCTGGGGAGGCTGGCTGGAGGGTGGGGTGTTGTGGTGGCTCCCTCCTCCTCCTCCTCCTCC-
EF0074784...    GCCTGTGGGCCTGGGGTATT--------------------------------------------------------------
EF0074785...    GCCTGTGGGCCTGGGGTATT--------------------------------------------------------------
EF0074785...    GCCTGTGGGCCTGGGGTATTTCTGGCTGGGGAGGCTGGCTGGAGGGTGGGGTGTTGTGGTGGCTCCCTCCTCCTCCTCCTCCTCCT
EF0074784...    ------------------------------------------------CTGGAGGGTGGGGTGTTGTGGTGGCTCCCTCCTCCTCCTCCTCCTCCT
```

```
                2611                                                                                                          2692
CgA-GLuc_...    cccaagggcctgattgtccctaccttggactcctgcatggggctccggaagggctgggggctggctccgtctgtggc
52BH88_47...    ------------------------------------------------------------------------------------
EF0074784...    ------------------------------------------------------------------------------------
EF0074785...    ------------------------------------------------------------------------------------
EF0074784...    CCCAGGGGGCCTGATTGTTCCCCTCAGCCTTGGACTCCTGCATGGGGCTCCGGAAGGGGCTGGGGGGCTGGCTCCGTCTGTGGC
EF0074785...    CCCAGGGGGCCTGATTGTTCCCCTCAGCCTTGGACTCCTGCATGGGGCTCCGGAAGGGGCTGGGGGGCTGGCTCCGTCTGTGGC
EF0074784...    CCCAGGGGGCCTGATTGTTCCCCTCAGCCTTGGACTCCTGCATGGGGCTCCGGAAGGGGCTGGGGGGCTGGCTCCGTCTGTGGC
```

```
                2693                                                                                                          2774
CgA-GLuc_...    ttcaccaacttttcttctgcctccttggaatcctctctttttctccatacaatccttggatgatggctccttccaacgctcttc
52BH88_47...    ------------------------------------------------------------------------------------------
EF0074784...    ------------------------------------------------------------------------------------------
EF0074785...    ------------------------------------------------------------------------------------------
EF0074784...    ------------------------------------------------------------------------------------------
EF0074785...    TTCACCAACTTTTCTTCTGCCTCCTTGGAATCCTCTCTTTTTCTCCATAAACATCCTTGGATGATGGCTCCTTCCACCGGCTCTTTC
EF0074785...    TTCACCAACTTTTCTTCTGCCTCCTTGGAATCCTCTCTTTTTCTCCATAAACATCCTTGGATGATGGCTCCTTCCACCGGCTCTTTC
EF0074784...    ------------------------------------------------------------------------------------------
```

```
                2775                                                                                                          2856
CgA-GLuc_...    agctcaggcctggctctcgttctcaagaacctgaggagttcatcttcaaaacccgctgtgttctctgttctgatgtgccc
52BH88_47...    ----------------------------------------------------------------------------------------
EF0074784...    ----------------------------------------------------------------------------------------
EF0074785...    ----------------------------------------------------------------------------------------
EF0074784...    ----------------------------------------------------------------------------------------
EF0074785...    AGCTCGGCCTGGCTCTCGGTTCTCAAGAACCTGAGAGAGTTCATCTTCAAAAACCGCTGTGTTTCTTCTCGTGATGTGCCC
EF0074785...    AGCTCGGCCTGGCTCTCGGTTCTCAAGAACCTGAGAGAGTTCATCTTCAAAAACCGCTGTGTTTCTTCTCGTGATGTGCCC
EF0074784...    AGCTCGGCCTGGCTGCTCTGGTTCTCAAGAACCTGAGAGAGTCATCTTCAAAAACCCGCTGTGTTTCTTCTCGTGATGTGCCC
```

Section 6.4

Figure 118: Sanger sequencing alignments for donor plasmid with Ef1-HTLV promoter and CgA-GLuc. Any mismatched bases are indicated in red (5 pages).

6.4.8 Donor plasmid with Ef1-HTLV promoter and SgII-GLuc

Sanger sequencing alignment for SgII-GLuc in Ef1-HTLV promoter plasmid. EF0074785: Plasmid 1, forward sequencing run; EF0074784: Plasmid 3, forward sequencing run; 52BH87: Plasmid 1, forward sequencing run, part 2; EF0074784_2: Plasmid 3, forward sequencing run, part 2; 52BH86: Plasmid 1, middle sequencing run; EF0074785_2: Plasmid 3, middle sequencing run; 52BH85: Plasmid 1, reverse sequencing run; EF0074784_3: Plasmid 3, reverse sequencing run (6 pages).

Page 1 EF00747850_EF00747850 (EF00747850_EF00747850.ab1), EF00747846_EF00747846 (EF00747846...

800
SgII-Gluc... ctagtcacacggccccctgatctgtccaacctggcctggatctgtctgcgaaaggtgcacaagcttgcggcagcac
EF0074785... CTAGTCACACCGGCGCCCCTGATCTGTGTCCAACCTGGCCCTGGATCTTGTGTCGCAAAGGTCGCACACGGTTGGGCAGCCAC
EF0074784... CTAGTCACACCGGCGCCCCTGATCTGTGTCCAACCTGGCCCTGGATCTTGTGTCGCAAAGGTCGCACACGGTTGGGCAGCCAC
EF0074784... CTAGTCACACCGGCGCCCCTGATCTGTGTCCAACCTGGCCCTGGATCTTGTGTCGCAAAGGTCGCACACGGTTGGGCAGCCAC
5ZBH87_47... ----------------------------CCTGGATCTGTGTCGGCAAAGGTCGCACACGGTTGGGCAGCCAC
5ZBH87_47... ----------------------------CCTGGATCTGTGTCGGCAAAGGTCGCACACGGTTGGGCAGCCAC
EF0074784...
EF0074784...
5ZBH86_47...
5ZBH86_47...
EF0074785...
EF0074785...
5ZBH85_47...
5ZBH85_47...
EF0074784...

882
SgII-Gluc... tcttgagcaggtcagaaaactgcacgttggcaaagcccttgaggcagcagttgtcgagtccacacacagatcgaccgtg
EF0074785... TTCTTGAGCAGGTCAGAACACTGCACGTTGGCAGCGCCAGCCAGTTGTCGAGTCCACACACAGATCGACCGTGTG
EF0074784... TTCTTGAGCAGGTCAGAACACTGCACGTTGGCAGCGCCAGCCAGTTGTCGAGTCCACACACAGATCGACCGTGTG
EF0074784... TTCTTGAGCAGGTCAGAACACTGCACGTTGGCAGCGCCAGCCAGTTGTCGAGTCCACACACAGATCGACCGTGTG
5ZBH87_47... TTCTTGAGCAGGTCAGAACACTGCACGTTGGCAGCGCCAGCCAGTTGTCGAGTCCACACACAGATCGACCGTGTG
5ZBH87_47... TTCTTGAGCAGGTCAGAACACTGCACGTTGGCAGCGCCAGCCAGTTGTCGAGTCCACACACAGATCGACCGTGTG
EF0074784...
EF0074784...
5ZBH86_47...
5ZBH86_47...
EF0074785...
EF0074785...
5ZBH85_47...
5ZBH85_47...
EF0074784...

964
SgII-Gluc... cgatgaactgctccatgggctccaagtccttgaaccaggaatctcaggaaatgtcgacgacgtgctcgcctatgcccgcctg
EF0074785... CGATGAACTGCTCCATGGGCTCCAAGTCCTTGAACCAGGAATCTCAGGAAATGTCGACGACATCGACCTCGCCTATGCCCGCCCTG
EF0074784... CGATGAACTGCTCCATGGGCTCCAAGTCCTTGAACCAGGAATCTCAGGAAATGTCGACGACATCGACCTCGCCTATGCCCGCCCTG
EF0074784... CGATGAACTGCTCCATGGGCTCCAAGTCCTTGAACCAGGAATCTCAGGAAATGTCGACGACATCGACCTCGCCTATGCCCGCCCTG
5ZBH87_47... CGATGAACTGCTCCATGGGCTCCAAGTCCTTGAACCAGGAATCTCAGGAAATGTCGACGACATCGACCTCGCCTATGCCCGCCCTG
5ZBH87_47... CGATGAACTGCTCCATGGGCTCCAAGTCCTTGAACCAGGAATCTCAGGAAATGTCGACGACATCGACCTCGCCTATGCCCGCCCTG
EF0074784...
EF0074784...
5ZBH86_47...
5ZBH86_47...
EF0074785...
EF0074785...
5ZBH85_47...
5ZBH85_47...
EF0074784...

1046
SgII-Gluc... tgcgggactcattgtccgccttcgtagtgtggcaggacgctgggaatgaactcttccatcttgaggcacctccagcactgatgtgggac
EF0074785... TGCGGGACTCTTTGTCTCGCCTTCGTAGGTGTGGCAGGACGTCGGGAATGAACTTCTTCCATCTTCAATCTTGAGCACCTCCAGTGATGTGGGAC
EF0074784... TGCGGGACTCTTTGTCTCGCCTTCGTAGGTGTGGCAGGACGTCGGGAATGAACTTCTTCCATCTTCAATCTTGAGCACCTCCAGTGATGTGGGAC
EF0074784... TGCGGGACTCTTTGTCTCGCCTTCGTAGGTGTGGCAGGACGTCGGGAATGAACTTCTTCCATCTTCAATCTTGAGCACCTCCAGTGATGTGGGAC
5ZBH87_47... TGCGGGACTCTTTGTCTCGCCTTCGTAGGTGTGGCAGGACGTCGGGAATGAACTTCTTCCATCTTCAATCTTGAGCACCTCCAGTGATGTGGGAC
5ZBH87_47... TGCGGGACTCTTTGTCTCGCCTTCGTAGGTGTGGCAGGACGTCGGGAATGAACTTCTTCCATCTTCAATCTTGAGCACCTCCAGTGATGTGGGAC
EF0074784...
EF0074784...
5ZBH86_47...
5ZBH86_47...
EF0074785...
EF0074785...
5ZBH85_47...
5ZBH85_47...
EF0074784...

1128
SgII-Gluc... aggcagagaatcagagacagccccctggtgcagccgaccagctttccgggcattggcttccaatctctcttgagcaacctccaggggcagctctt
EF0074785... AGGCAGATCAGACAGCACGCCCCTGGTGCAGCCGACCAGCTTTCCCGGGCATTGGCTTCCATCTCTCTTGAGCAGGGGCAGGCTTCT
EF0074784... AGGCAGATCAGACAGCACGCCCCTGGTGCAGCCGACCAGCTTTCCCGGGCATTGGCTTCCATCTCTCTTGAGCAGGGGCAGGCTTCT
EF0074784... AGGCAGATCAGACAGCACGCCCCTGGTGCAGCCGACCAGCTTTCCCGGGCATTGGCTTCCATCTCTCTTGAGCAGGGGCAGGCTTCT
5ZBH87_47... AGGCAGATCAGACAGCACGCCCCTGGTGCAGCCGACCAGCTTTCCCGGGCATTGGCTTCCATCTCTCTTGAGCAGGGGCAGGCTTCT
5ZBH87_47... AGGCAGATCAGACAGCACGCCCCTGGTGCAGCCGACCAGCTTTCCCGGGCATTGGCTTCCATCTCTCTTGAGCAGGGGCAGGCTTCT
EF0074784...
EF0074784...
5ZBH86_47...
5ZBH86_47...
EF0074785...
EF0074785...
5ZBH85_47...
5ZBH85_47...
EF0074784...

Page 2 EF00747850_EF00747850 (EF00747850_EF00747850.ab1), EF00747846_EF00747846 (EF00747846...

1210
SgII-Gluc... tgccggggcaactccccgcgtcagatcagatccggtgcggaagtgtctggccacggccacgacgtcgaagtcttcgtt
EF0074785... TGCCGGGCAACTTCCCCGCGGGTCAGACATCCAACATCGACATCGAGAGTCCGGCCGCCGGACCACGACGTTCGTT
EF0074784... TGCCGGGCAACTTCCCCGCGGGTCAGACATCCAACATCGAGAGTCCGGCCGCCGGACCACGACGTTCGTT
EF0074784... TGCCGGGCAACTTCCCCGCGGGTCAGACATCCAACATCGAGAGTCCGGCCGCCGGACCACGACGTTCGTT
5ZBH87_47... TCCCGGGCCACTTCCCGCGGGTCAGACATCGACATCCAGAGTCCGGCCGCCGGACCACGACGTTCGTT
5ZBH87_47... TCCCGGGCCACTTCCCGCGGGTCAGACATCGACATCCAGAGTCCGGCCGCCGGACCACGACGTTCGTT
EF0074784...
EF0074784...
5ZBH86_47...
5ZBH86_47...
EF0074785...
EF0074785...
5ZBH85_47...
5ZBH85_47...
EF0074784...

1292
SgII-Gluc... gttctcggtgggctcatcatcatttccatgtctcttagcaatatgctccctccctttctgcctttttcttgattgagg
EF0074785... GTTCTCGGTGGGCTCATCATATTTCCATTGCTCTCTTAGCAATAATGCTCCCTTCCCTTTTCTGCCTTTTTCTTCGTTGAGG
EF0074784... GTTCTCGGTGGGCTCATCATATTTCCATTGCTCTCTTAGCAATAATGCTCCCTTCCCTTTTCTGCCTTTTTCTTCGTTGAGG
EF0074784... GTTCTCGGTGGGCTCATCATATTTCCATTGCTCTCTTAGCAATAATGCTCCCTTCCCTTTTCTGCCTTTTTCTTCGTTGAGG
5ZBH87_47... CATTGCTCTCTTAGCAATAATGCTCCCTTCCCTTTTCTGCCTTTTTCTTCGTTGAGG
5ZBH87_47... CATTGCTCTCTTAGCAATAATGCTCCCTTCCCTTTTCTGCCTTTTTCTTCGTTGAGG
EF0074784...
EF0074784...
5ZBH86_47...
5ZBH86_47...
EF0074785...
EF0074785...
5ZBH85_47...
5ZBH85_47...
EF0074784...

1374
SgII-Gluc... tattccagcaacttcattaacagatccttcatcccagtactgctcattgggggtatcatcatcttcggggcccacagggga
EF0074785... TATTCCAGCACTTTCATTAACAGATCTTCATCCCAGTACTGCCTATTGGGGTATCATCATCTTCGGGGGGCCCACAGGGA
EF0074784... TATTCCAGCACTTTCATTAACAGATCTTCATCCCAGTACTGCCTATTGGGGTATCATCATCTTCGGGGGGCCCACAGGGA
EF0074784... TATTCCAGCACTTTCATTAACAGATCTTCATCCCAGTACTGCCTATTGGGGTATCATCATCTTCGGGGGGCCCACAGGGA
5ZBH87_47... TATTCCAGCACTTTCATTAACAGATCTTCATCCCAGTACTGCCTATTGGGGTATCATCATCTTCGGGGGGCCCACAGGGA
5ZBH87_47... TATTCCAGCACTTTCATTAACAGATCTTCATCCCAGTACTGCCTATTGGGGTATCATCATCTTCGGGGGGCCCACAGGGA
EF0074784...
EF0074784...
5ZBH86_47...
5ZBH86_47...
EF0074785...
EF0074785...
5ZBH85_47...
5ZBH85_47...
EF0074784...

1456
SgII-Gluc... acctttgctcaccggggccagctgtcagtctcctgagagctgccttgattcaaatgctcttgatggctgctcaattg
EF0074785... ACCTTTGCTCACCGGGGCCAGCTGTCAGTCTCCTGAGAGCTGCCTTGATTCAAATGCTCTTTGATGGCCTGCTCAATTTG
EF0074784... ACCTTTGCTCACCGGGGCCAGCTGTCAGTCTCCTGAGAGCTGCCTTGATTCAAATGCTCTTTGATGGCCTGCTCAATTTG
EF0074784... ACCTTTGCTCACCGGGGCCAGCTGTCAGTCTCCTGAGAGCTGCCTTGATTCAAATGCTCTTTGATGGCCTGCTCAATTTG
5ZBH87_47... ACCTTTGCTCACCGGGGCCAGCTGTCAGTCTCCTGAGAGCTGCCTTGATTCAAATGCTCTTTGATGGCCTGCTCAATTTG
5ZBH87_47... ACCTTTGCTCACCGGGGCCAGCTGTCAGTCTCCTGAGAGCTGCCTTGATTCAAATGCTCTTTGATGGCCTGCTCAATTTG
EF0074784...
EF0074784...
5ZBH86_47... ACCTTTTGCTCACCGGGGCCAGCTTGTCAGTCTCCTGAGAGCTGCCTTGATTCAAATGCTCTTTGATGGCCTGCTCAATTTG
5ZBH86_47... ACCTTTTGCTCACCGGGGCCAGCTTGTCAGTCTCCTGAGAGCTGCCTTGATTCAAATGCTCTTTGATGGCCTGCTCAATTTG
EF0074785...
EF0074785...
5ZBH85_47... TTCCTCTTCTTCCTGCAGGTCATCTTCAGATGAGCCTTGACCAGGAGAACTCGCTTCACTTGGTTTGAATTAATGATCTCAGGGTAT
5ZBH85_47... TTCCTCTTCTTCCTGCAGGTCATCTTCAGATGAGCCTTGACCAGGAGAACTCGCTTCACTTGGTTTGAATTAATGATCTCAGGGTAT
EF0074784...

```
                1620                                                                                                              1701
SgII-GLuc...    ttaactagcatcctggcaagtactcaactaattccttgatccttgtcgttcaggtttcatatgcaatcctgtctgtttcaa
EF0074785...    --------------------------------------------------------------------------------
EF0074784...    --------------------------------------------------------------------------------
52BH87_47...    TTAACTAGCATCCTGGCCAAGTACTCAACTCACCTAATTCTTGATCCCTTGTCGTTCAGGTTTCATATATGCCATCTGTCTGTTTCAA
EF0074784...    --------------------------------------------------------------------------------
52BH86_47...    TTAACTAGCATCCTGGCCAAGTACTCAACTCACCTAATTCTTGATCCCTTGTCGTTCAGGTTTCATATATGCCATCTGTCTGTTTCAA
EF0074785...    TTAACTAGCATCCTGGCCAAGTACTCAACTAATTCTTGATCCCTTGTCGTTCAGGTTTCATATATGCCATCTGTCTGTTTCAA
52BH85_47...    TTAACTAGCATCCTGGCCAAGTACTCAACTAATTCTTGATCCTTGTCGTTCAGGTTTCATATATGCCATCTGTCTGTTTCAA
EF0074784...    --------------------------------------------------------------------------------

                ....................................................................................................

                1702                                                                                                              1783
SgII-GLuc...    catgttgagaatccaggcagcttgggaagctggttcgatctagatcttccagcacacataagggagccttggcagaaacttctc
EF0074785...    --------------------------------------------------------------------------------
EF0074784...    --------------------------------------------------------------------------------
52BH87_47...    CATGTGGAATCCAGGCAGCAGCTTTGGGAAGCTTCGAATCTAGATCTTCCAGCACCATAAGGGAGCCTTGGCAGAACTTCTC
EF0074784...    --------------------------------------------------------------------------------
52BH86_47...    CATGTGGAATCCAGGCAGCCAGCTTTGGGAAGCTGGTTCGAATCTAGATCTTCCAGCACCATAAGGGAGCCTTGGCAGAACTTCTC
EF0074785...    CATGTGGAATCCAGGCAGCCAGCTTTGGGAAGCTGGTTCGAATCTAGATCTTCCAGCACCATAAGGGAGCCTTGGCAGAACTTCTC
52BH85_47...    CATGTGGAATCCAGGCAGCCAGCTTTGGGAAGCTGGTTCGAATCTAGATCTTCCAGCACCATAAGGGAGCCTTGGCAGAACTTCTC
EF0074784...    --------------------------------------------------------------------------------

                ....................................................................................................

                1784                                                                                                              1865
SgII-GLuc...    ctggttatatggattaggaaaatacgacagtttctgatttgctgcaactctccatccctaaaagatttaaaatatcctcaaca
EF0074785...    --------------------------------------------------------------------------------
EF0074784...    --------------------------------------------------------------------------------
52BH87_47...    CTGGTTATATGGATTAGGAAAATACGACAGTTTTCTGATTTGCTGCACTCTCCATCCCTAAAAGATTTAAAATATCCTCAACA
EF0074784...    --------------------------------------------------------------------------------
52BH86_47...    CTGGTTATATGGATTGGGGAAAAATACGACAGTTTTCTGATTTGCTGCACTCTCCATCCCTAAAAAGATTTAAAATATCCTCAACA
EF0074785...    CTGGTTATATGGATTGGGAAAATACGACGTTTTCTCGATTTCTGCATTCCTCGCACTCCCATCCCTAAAAGATTTAAAATATCCTCAACA
52BH85_47...    --------------------------------------------------------------------------------
EF0074784...    --------------------------------------------------------------------------------

                ....................................................................................................

                1866                                                                                                              1947
SgII-GLuc...    ctgagcccgctcggtaggcctccagtcccagacacaggtgttcttaaggtagccactcttggagaggcatcctatttgga
EF0074785...    --------------------------------------------------------------------------------
EF0074784...    --------------------------------------------------------------------------------
52BH87_47...    CTGAGCCCGTCTCGGTAGGGCCTCCAGTCCCAGCACACGACCAGGTGTTTTAGGGTAGCCACTCTTGGAGAGCATCCTATTTTGGA
EF0074784...    --------------------------------------------------------------------------------
52BH86_47...    CTGAGCCCGTCTCGGTAGGGCCTCAGTCCCAGCACGACCAGGTGTTTTAGGGTAGCCACTCTTGGAGAGCATCCTATTTTGGA
EF0074785...    CTGAGCCCGTC-------------------------------------------------------------------
52BH85_47...    --------------------------------------------------------------------------------
EF0074784...    --------------------------------------------------------------------------------

                ....................................................................................................

                1948                                                                                                              2029
SgII-GLuc...    acaagtctggatggtcttaagtcagcctctgagatgtcatctaggtcaacagaagaggtcaagctccgctccggttccactga
EF0074785...    --------------------------------------------------------------------------------
EF0074784...    --------------------------------------------------------------------------------
52BH87_47...    ACAGGTCTGGATGGTCTAAGTCAGCCTCTGA-------------------------------------------------
EF0074784...    --------------------------------------------------------------------------------
52BH86_47...    ACAGGTCTGGATGGTCTAAGTCAGCTCTCGAGAGTCATCTAGGTCAACAGAGAGAGGTCAAGCTCCGGTCCGGTTCCACTGA
EF0074785...    --------------------------------------------------------------------------------
52BH85_47...    --------------------------------------------------------------------------------
EF0074784...    --------------------------------------------------------------------------------

                ....................................................................................................
```

```
                2030                                                                                                              2111
SgII-GLuc...    tccattccggcttctcccccagtttgagcatctcaattaagtctctgggggtatctgtaaattccttgagattccaatcagc
EF0074785...    --------------------------------------------------------------------------------
EF0074784...    --------------------------------------------------------------------------------
52BH87_47...    --------------------------------------------------------------------------------
EF0074784...    --------------------------------------------------------------------------------
52BH86_47...    TCCATTCGGCCTTCTCCCCAGTTTTGAGCATCTCAATTAAGTCTCTGGGGTATCTGTAAATTCCTTGAGATTTCAATCAGC
EF0074785...    --------------------------------------------------------------------------------
52BH85_47...    --------------------------------------------------------------------------------
EF0074784...    --------------------------------------------------------------------------------

                ....................................................................................................

                2112                                                                                                              2193
SgII-GLuc...    tgataaatagactgagaatcaagagggttctctcaaaaaagcctggtggccctttccccatttgcccatcttgtaacctcccac
EF0074785...    --------------------------------------------------------------------------------
EF0074784...    --------------------------------------------------------------------------------
52BH87_47...    --------------------------------------------------------------------------------
EF0074784...    --------------------------------------------------------------------------------
52BH86_47...    TGATAAATAGACTGAGAATCAAGAGAGGTTTCTCAAAAAAGCCTGGTGGCCCTTTCCCCATTTGCCCATTCTTGTAACCTCCCAC
EF0074785...    --------------------------------------------------------------------------------
52BH85_47...    ------------------------------------------------------------------GTAACCTCCCAC
EF0074784...    --------------------------------------------------------------------------------

                ....................................................................................................

                2194                                                                                                              2275
SgII-GLuc...    ttcctgcagcattactaacctcttcaaataggcaattactacttggagaacatcatctgagagttgttcttacctcttccg
EF0074785...    --------------------------------------------------------------------------------
EF0074784...    --------------------------------------------------------------------------------
52BH87_47...    --------------------------------------------------------------------------------
EF0074784...    --------------------------------------------------------------------------------
52BH86_47...    TTCCTGCAGCATTACTAACCTTTCAAATAGGCAATTACTTTGGAGAACATCATCTGAGAGTTGGTCTTACCTCTTCCG
EF0074785...    --------------------------------------------------------------------------------
52BH85_47...    TTCCTGCAGCATTACTAACCTTTCAAATAGGCAATTACTTCTGGAGACATCATCTGAGAGTTGGTCTTACTCTCTTCCG
EF0074784...    --------------------------------------------------------------------------------

                ....................................................................................................

                2276                                                                                                              2357
SgII-GLuc...    aagatctcttcctggatgccaagctgccctgagcgttcatctcatcgtgattgtcaatttttcctatattcctcttg
EF0074785...    --------------------------------------------------------------------------------
EF0074784...    --------------------------------------------------------------------------------
52BH87_47...    --------------------------------------------------------------------------------
EF0074784...    --------------------------------------------------------------------------------
52BH86_47...    AAGATCTCTTCCTGGATGCCAAGCTGCCCTGAGCGTTCATCTCATCGTGATTGTTGTTCATTTTTTTCCTATATTCCTTG
EF0074785...    AAGATCTCTTCCTGGATGCCAAGCTGCCCTGAGCGTTCATCTCATCGTGATTGTTGTTCATTTTTTCATTTTTTCTCTTTG
52BH85_47...    AAGATCTCTTCTCCTGGATGCCAAGCTGCCCTGAGCGTTCATCTCATCGTGATTGTTGTTCATTTTTTTCATTTTTTCTCTTTG
EF0074784...    --------------------------------------------------------------------------------

                ....................................................................................................

                2358                                                                                                              2439
SgII-GLuc...    ctgtctctcacctcttcctgggtttgactctctatttctcctctactgggttccagtcttctcccccgacacacatcctcat
EF0074785...    --------------------------------------------------------------------------------
EF0074784...    --------------------------------------------------------------------------------
52BH87_47...    --------------------------------------------------------------------------------
EF0074784...    --------------------------------------------------------------------------------
52BH86_47...    CTGTCTCTCACCTCTTCCTGGGTTTGACTCTCTATTTTCTCCTCTACTGGGTTCCCCGACACACATCTCAT
EF0074785...    CTGTCTCTCACCTCTTCCTGGGTTTGACTCTCTATTTTCTCTCTACTGGGTTCCCCGACACACATCTCAT
52BH85_47...    CTGTCTCTCACCTCTTCCTGGGTTTGACTCTCTATTTTCTCCTCTACTGGGTTCCCCGACACACATCTTCAT
EF0074784...    --------------------------------------------------------------------------------

                ....................................................................................................
```

Section 6.4

Figure 119: Sanger sequencing alignments for donor plasmid with Ef1-HTLV promoter and SgII-GLuc. Any mismatched bases are indicated in red (6 pages).

6.4.9 Donor plasmid with Ef1-HTLV promoter and VAMP2-GLuc

Sanger sequencing alignment for VAMP2-GLuc in Ef1-HTLV promoter plasmid. CQA940: Plasmid 1, forward sequencing run; CQA945: Plasmid 1, reverse sequencing run (2 pages).

Figure 120: Sanger sequencing alignments for donor plasmid with Ef1-HTLV promoter and VAMP2-GLuc. Any mismatched bases are indicated in red (2 pages).

Section 6.4

6.4.10 IMR90-4 hPOMC-GLuc clone 6

Sanger sequencing alignment for IMR90-4 hPOMC-GLuc (6) donor DNA insertion. Traces corresponding to dark red sequencing arrows depicted in Figure 64 in the Results section Sanger sequencing of homozygote clones: 52BH19: sequencing range #1, 52BH20: sequencing range #2, 52BH21: sequencing range #3, 52BH08: sequencing range #4, 52BH62: sequencing range #5, 52BH07: sequencing range #6, 52BH61: sequencing range #7, 52BH63: sequencing range #8, 43EH26: sequencing range #9, 19HI98: sequencing range #10 (12 pages).

```
                    1                                                              82
F10_seq           TTCAGGTTCCGTCTTCCTCCACTCCCTCTTCCCCTTGCTCTCTGCTGTGTTGCTGACCCAAGATAGCTCTTCCGGACACTT
52BH20_46...      TTCAGGTTCCGTCTTCCTCCACTCCCTCTTCCCCTTGCTCTCTGCTGTGTTGCTGACCCAAGATAGCTCTTCCGGACACTT
52BH20_46...      ------------------TCCCTCTTCCCCTTGCTCTCTGCTGTGTTGCTGACCCAAGATAGCTCTTCCGGACACTT
52BH21_46...      -------------------TCCCTCTTCCCCTTGCTCTCTGCTGTGTTGCTGACCCAAGATAGCTCTTCCGGACTGTGGAT
52BH19_46...      -------------------------------------GCTCTTCTTCCGGACACTT
43EH26_(4...
43EH26_(4...
52BH07_46...
52BH08_46...
52BH07_46...
52BH63_47...
52BH61_47...
52BH61_47...
19H198_04...
52BH62_47...

                    83                                                             164
F10_seq           CCTTCTCGGCGCCTGCACCACGTGATCGTCTCTGAGCGGAGATCCTCCCCGGTGCTGGGTGCCTCCCGGCGCATCTCTCCCTC
52BH20_46...      CCTTCTCGGCGCCTGCACCACGTGATCGTCTCTGAGCGGAGATCCTCCCCGGTGCTGGGTGCCTCCCGGCGCATCTCTCCCTC
52BH20_46...      CCTTCTCGGCGCCTGCACCACGTGATCGTCTCTGAGCGGAGATCCTCCCCGGTGCTGGGTGCCTCCCGGCGCATCTCTCCCTC
52BH21_46...      CCTTCTCGGCGCCTGCACCACGTGATCGTCTCTGAGCGGAGATCCTCCCCGGTGCTGGGTGCCTCCCGGCGCATCTCTCCCTC
52BH19_46...      CCTTCTCGGCGCCTGCACCACGTGATCGTCTCTGAGCGGAGATCCTCCCCGGTGCTGGGTGCCTCCCGGCGCATCTCTCCCTC
43EH26_(4...
43EH26_(4...
52BH07_46...
52BH08_46...
52BH07_46...
52BH63_47...
52BH61_47...
52BH61_47...
19H198_04...
52BH62_47...

                    165                                                            246
F10_seq           ACCCAACCCCATGCCGTCTTCACTCCGCTGGGTTCCCTTTTCCTTCTCTCCTTCTCGGGGGCCGTGCCCATCTCTCGGTTTCTTAGGA
52BH20_46...      ACCCAACCCCATGCCGTCTTCACTCCGCTGGGTTCCCTTTTCCTTCTCTCCTTCTCGGGGGCCGTGCCCATCTCTCGGTTTCTTAGGA
52BH21_46...      ACCCAACCCCATGCCGTCTTCACTCCGCTGGGTTCCCTTTTCCTTCTCTCCTTCTCGGGGGCCGTGCCCATCTCTCGGTTTCTTAGGA
52BH19_46...      ACCCAACCCCATGCCGTCTTCACTCCGCTGGGTTCCCTTTTCCTTCTCTCCTTCTCGGGGGCCGTGCCCATCTCTCGGTTTCTTAGGA
43EH26_(4...      ACCCAACCCCATGCCGTCTTCACTCCGCTGGGTTCCCTTTTCCTTCTCTCCTTCTCGGGGGCCGTGCCCATCTCTCGGTTTCTTAGGA
43EH26_(4...
52BH07_46...
52BH08_46...
52BH07_46...
52BH63_47...
52BH61_47...
52BH61_47...
19H198_04...
52BH62_47...

                    247                                                            328
F10_seq           TGGCCTTCTCCGACGGATGTCTCCCTTGGGTCCCGGCCTCCCCCCTTCCTTGTAGGCCTCGCATCACACCGTTTTCTGGACAACC
52BH20_46...      TGGCCTTCTCCGACGGATGTCTCCCTTGGGTCCCGGCCTCCCCCCTTCCTTGTAGGCCTCGCATCACACCGTTTTCTGGACAACC
52BH20_46...      TGGCCTTCTCCGACGGATGTCTCCCTTGGGTCCCGGCCTCCCCCCTTCCTTGTAGGCCTCGCATCACACCGTTTTCTGGACAACC
52BH21_46...      TGGCCTTCTCCGACGGATGTCTCCCTTGGGTCCCGGCCTCCCCCCTTCCTTGTAGGCCTCGCATCACACCGTTTTCTGGACAACC
52BH19_46...      TGGCCTTCTCCGACGGATGTCTCCCTTGGGTCCCGGCCTCCCCCCTTCCTTGTAGGCCTCGCATCACACCGTTTTCTGGACAACC
43EH26_(4...      TGGCCTTCTCCGACGGATGTCTCCCTTGGGTCCCGGCCTCCCCCCTTCCTTGTAGGCCTCGCATCACACCGTTTTCTGGACAACC
43EH26_(4...
52BH07_46...
52BH08_46...
52BH07_46...
52BH63_47...
52BH61_47...
52BH61_47...
19H198_04...
52BH62_47...
```

```
                    329                                                            410
F10_seq           CCAAAGTACCCCGTCTCCCTGGCTTTAGCCACCTCTCCATCTCTCTTGCTTCTTCTTGCCTCGGACACCCCGTTCTCCGTGTGGAT
52BH20_46...      CCAAAGTACCCCGTCTCCCTGGCTTTAGCCACCTCTCCATCTCTCTTGCTTCTTCTTGCCTCGGACACCCCGTTCTCCGTGTGGAT
52BH21_46...      CCAAAGTACCCCGTCTCCCTGGCTTTAGCCACCTCTCCATCTCTCTTGCTTCTTCTTGCCTCGGACACCCCGTTCTCCGTGTGGAT
52BH19_46...      CCAAAGTACCCCGTCTCCCTGGCTTTAGCCACCTCTCCATCTCTCTTGCTTCTTCTTGCCTCGGACACCCCGTTCTCCGTGTGGAT
43EH26_(4...      CCAAAGTACCCCGTCTCCCTGGCTTTAGCCACCTCTCCATCTCTCTTGCTTCTTCTTGCCTCGGACACCCCGTTCTCCGTGTGGAT
43EH26_(4...
52BH07_46...
52BH08_46...
52BH07_46...
52BH63_47...
52BH61_47...
52BH61_47...
19H198_04...
52BH62_47...

                    411                                                            492
F10_seq           TCGGGTCACCTCTCACTCCTTTCATTTGGGCAGCTCCCCTACCCCCCTTACCTCTCTAGTCTGTGCTAGTCTTCCAGCCCC
52BH20_46...      TCGGGTCACCTCTCACTCCTTTCATTTGGGCAGCTCCCCTACCCCCCTTACCTCTCTAGTCTGTGCTAGTCTTCCAGCCCC
52BH21_46...      TCGGGTCACCTCTCACTCCTTTCATTTGGGCAGCTCCCCTACCCCCCTTACCTCTCTAGTCTGTGCTAGTCTTCCAGCCCC
52BH19_46...      TCGGGTCACCTCTCACTCCTTTCATTTGGGCAGCTCCCCTACCCCCCTTACCTCTCTAGTCTGTGCTAGTCTTCCAGCCCC
43EH26_(4...      TCGGGTCACCTCTCACTCCTTTCATTTGGGCAGCTCCCCTACCCCCCTTACCTCTCTAGTCTGTGCTAGTCTTCCAGCCCC
43EH26_(4...
52BH07_46...
52BH08_46...
52BH07_46...
52BH63_47...
52BH61_47...
52BH61_47...
19H198_04...
52BH62_47...

                    493                                                            574
F10_seq           CTGTCAATGGCATCTTCCAGGGGTCCGAGAGCTCAGCTAGTCTTCTTCCTCCAACCCGGGGCCCCTATGTCCACTTCAGGACAG
52BH20_46...      CTGTCAATGGC--------------------
52BH21_46...      CTGTCAATGGCATCTTCCAGGGGTCCGAGAGCTCAGCTAGTCTTCTTCCTCCAACCCGGGGCCCCTATGTCCACTTCAGGACAG
52BH19_46...      CTGTCAATGGCATCTTCCAGGGGTCCGAGAGCTCAGCTAGTCTTCTTCCTCCAACCCGGGGCCCCTATGTCCACTTCAGGACAG
43EH26_(4...      CTGTCAATGGCATCTTCCAGGGGTCCGAGAGCTCAGCTAGTCTTCTTCCTCCAACCCGGGGCCCCTATGTCCACTTCAGGACAG
43EH26_(4...
52BH07_46...
52BH08_46...
52BH07_46...
52BH63_47...
52BH61_47...
52BH61_47...
19H198_04...
52BH62_47...

                    575                                                            656
F10_seq           CATCTTTGCTGCCTCCAGGGATCCTGTGTCCCCGAGGTGGGACACCTTATATATCCCAGGGGCCGGTAAGTGTGGCTCTGGTT
52BH20_46...      CATGTTTGCTGCCTCCAGGGATCCTGTGTCCCCGAGGTGGGACACCTTATATATCCCAGGGGCCGGTAAGTGTGGCTCTGGTT
52BH21_46...      CATGTTTGCTGCCTCCAGGGATCCTGTGTCCCCGAGGTGGGACACCTTATATATCCCAGGGGCCGGTAAGTGTGGCTCTGGTT
52BH19_46...      CATGTTTGCTGCCTCCAGGGATCCTGTGTCCCCGAGGTGGGACACCTTATATATCCCAGGGGCCGGTAAGTGTGGCTCTGGTT
43EH26_(4...      CATGTTTGCTGCCTCCAGGGATCCTGTGTCCCCGAGGTGGGACACCTTATATATCCCAGGGGCCGGTAAGTGTGGCTCTGGTT
43EH26_(4...
52BH07_46...
52BH08_46...
52BH07_46...
52BH63_47...
52BH61_47...
52BH61_47...
19H198_04...
52BH62_47...
```

Section 6.4

```
                657
F10_seq         CTGGGTACTTTTATCTGTCCCCTCCACCCCACCCACCCAGTGGGGCAGCAAGCTTCTTGACCTCTTCTTCTCTCTCCCCACAGGGGCCTCGAGAG   738
52BH20_46..
52BH21_46..     CTGGGTACTTTTATCTGTCCCCTCCACCCCACCCACCCAGTGGGGCAGCAAGCTTCTTGACCTCTTCTTCTCTCTCCCCACAGGGGCCTCGAGAG
52BH19_46..     CTGGGTACTTTTATCTGTCCCCTCCACCCCACCCACCCAGTGGGGCAGCAAGCTTCTTGACCTCTTCTTCTCTCTCCCCACAGGGGCCTCGAGAG
43EH26 (4.
52BH26 (4.      CTGGGTACTTTTATCTGTCCCCTCCACCCCACCCACCCAGTGGGGCAGCAAGCTTCTTGACCTCTTCTTCTCTCTCCCCACAGGGGCCTCGAGAG
52BH07_46..
52BH08_46..
52BH63_47..
52BH61_47..
19HI98_04..
52BH62_47..

                739
F10_seq         ATCTGGCACGCGGAGGAGGGCAAGAGAAGTCTTCTAACATGCGGTGACGTGGAGGAGGAGAATCCCGGCCCCTAGGCTCGAGATGACC   820
52BH20_46..
52BH21_46..     ATCTGGCACGCGGAGGAGGGCAAGAGAAGTCTTCTAACATGCGGTGACGTGGAGGAGGAGAATCCCGGCCCCTAGGCTCGAGATGACC
52BH19_46..     ATCTGGCACGCGGAGGAGGGCAAGAGAAGTCTTCTAACATGCGGTGACGTGGAGGAGGAGAATCCCGGCCCCTAGGCTCGAGATGACC
43EH26 (4.      ATCTGGCACGCGGAGGAGGGCAAGAGAAGTCTTCTAACATGCGGTGACGTGGAGGAGGAGAATCCCGGCCCCTAGGCTCGAGATGACC
52BH26 (4.
52BH07_46..
52BH08_46..
52BH63_47..
52BH61_47..
19HI98_04..
52BH62_47..

                821
F10_seq         GAGTACAAGCCCACCGGTGCCCTCCCCACCCCGCGACGAGCGTCCCCAGGGCCGTACGCACCCTCCCCGCCGCCGCCTTCGCCGACT   902
52BH20_46..
52BH21_46..     GAGTACAAGCCCACCGGTGCCCTCCCCACCCCGCGACGAGCGTCCCCAGGGCCGTACGCACCCTCCCCGCCGCCGCCTTCGCCGACT
52BH19_46..     GAGTACAAGCCCACCGGTGCCCTCCCCACCCCGCGACGAGCGTCCCCAGGGCCGTACGCACCCTCCCCGCCGCCGCCTTCGCCGACT
43EH26 (4.      GAGTACAAGCCCACCGGTGCCCTCCCCACCCCGCGACGAGCGTCCCCAGGGCCGTACGCACCCTCCCCGCCGCCGCCTTCGCCGACT
52BH26 (4.
52BH07_46..
52BH08_46..
52BH63_47..
52BH61_47..
19HI98_04..
52BH62_47..

                903
F10_seq         ACCCCGCCACGCGCCACACCGGTCGATCCGGACCGCACATCGAGGCGGGTCACCGAGCTGCAAGAGAACTCTTCCTCACCGCCGCGT   984
52BH20_46..
52BH21_46..     ACCCCGCCACGCGCCAC---
52BH19_46..     ACCCCGCCACCGCGCCACCGTCGATCCGGACCGCC-
43EH26 (4.      ACCCCGCCACCGCGCCACCGTCGATCCGGACCGCC-
52BH26 (4.
52BH07_46..
52BH08_46..
52BH63_47..
52BH61_47..
19HI98_04..
52BH62_47..
```

```
                985
F10_seq         CGGGCTCGACATCGGCACAAGGTGTGGGTCGCGGACGACGGCGGCCGCCGGGGTGCCGGTCTTGGACCACGGCCGGAGAGGCGTCGAAGCG   1066
52BH20_46..
52BH21_46..     CGGGCTCGACATCGGCACAAGGTGTGGGTCGCGGACGACGGCGGCCGCCGGGGTGCCGGTCTTGGACCACGGCCGGAGAGGCGTCGAAGCG
52BH19_46..
52BH26 (4.
43EH26 (4.
52BH07_46..
52BH08_46..
52BH63_47..
52BH61_47..
19HI98_04..
52BH62_47..

                1067
F10_seq         GGGGCCGGTGTTCGCCGAGATCGGCCCCGCGCATGGCCGAGTTGAGGGTTCCCGGCTGGCCGCCGCGACCAACAGATGGAAGGCC   1148
52BH20_46..
52BH21_46..     GGGGCCGGTGTTCGCCGAGATCGGCCCCGCGCATGGCCGAGTTGAGGGTTCCCGGCTGGCCGCCGCGACCAACAGATGGAAGGCC
52BH19_46..
52BH26 (4.
43EH26 (4.
52BH07_46..
52BH08_46..
52BH63_47..
52BH61_47..
19HI98_04..
52BH62_47..

                1149
F10_seq         TCCTGGCGCCCGCACCGGCCCCAAGGAGCCCCGCGTGGTTCCTGGCCACCGTCGGCGGTCTCGCCCGACCACCAGGGCCAAGGGTCT   1230
52BH20_46..
52BH21_46..     TCCTGGCGCCCGCACCGGCCCCAAGGAGCCCCGCGTGGTTCCTGGCCACCGTCGGCGGTCTCGCCCGACCACCAGGGCCAAGGGTCT
52BH19_46..
52BH26 (4.
43EH26 (4.
52BH07_46..
52BH08_46..
52BH63_47..
52BH61_47..
19HI98_04..
52BH62_47..

                1231
F10_seq         GGGCAGCGCGCCGTCGTGCTCCCCGGAGTGGAGGCGGCCGACCGCGCCGGGGTGCCCGCCCTTCCTGGAGACTTCCGCGCCCCGC   1312
52BH20_46..
52BH21_46..     GGGCAGCGCGCCGTCGTGCTCCCCGGAGTGGAGGCGGCCGACCGCGCCGGGGTGCCCGCCCTTCCTGGAGACTTCCGCGCCCCGC
52BH19_46..
52BH26 (4.
43EH26 (4.
52BH07_46..
52BH08_46..
52BH63_47..
52BH61_47..
19HI98_04..
52BH62_47..
```

229

Section 6.4

```
                    1313                                                                                    1394
F10.seq             AACCTCCCTTCTTACGAGCGGGCTCGGCTTCACCGTCACCCGCACGTCGAGGTGCCCGAAGGACCCGCGCACCTCGGTTGCATGA    1722
52BH20_46...        ----------------------------------------------------------------------------------
52BH21_46...        ----------------------------------------------------------------------------------
52BH19_46...        ----------------------------------------------------------------------------------
43EH26_(4...        ----------------------------------------------------------------------------------
52BH07_46...        ----------------------------------------------------------------------------------
52BH08_46...        ----------------------------------------------------------------------------------
52BH63_47...        ----------------------------------------------------------------------------------
52BH61_47...        ----------------------------------------------------------------------------------
19H198_04...        ----------------------------------------------------------------------------------
52BH62_47...        ----------------------------------------------------------------------------------

                    1395                                                                                    1476
F10.seq             CCCCCAAGCCCCGGTGCCTCGATCTAGAGGGCCCGTTTAAACCCCTCGATCACCCTCCACTGTTGCCTTCTTAGTTGCCAGCCATC
52BH20_46...        ----------------------------------------------------------------------------------
52BH21_46...        ----------------------------------------------------------------------------------
52BH19_46...        ----------------------------------------------------------------------------------
43EH26_(4...        ----------------------------------------------------------------------------------
52BH07_46...        ----------------------------------------------------------------------------------
52BH08_46...        ----------------------------------------------------------------------------------
52BH63_47...        ----------------------------------------------------------------------------------
52BH61_47...        ----------------------------------------------------------------------------------
19H198_04...        ----------------------------------------------------------------------------------
52BH62_47...        ----------------------------------------------------------------------------------

                    1477                                                                                    1558
F10.seq             TGTTGTTTCGCCCCTCCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTTAATAAAATGAGGAAAATT
52BH20_46...        ----------------------------------------------------------------------------------
52BH21_46...        ----------------------------------------------------------------------------------
52BH19_46...        ----------------------------------------------------------------------------------
43EH26_(4...        ----------------------------------------------------------------------------------
52BH07_46...        ----------------------------------------------------------------------------------
52BH08_46...        ----------------------------------------------------------------------------------
52BH63_47...        ----------------------------------------------------------------------------------
52BH61_47...        ----------------------------------------------------------------------------------
19H198_04...        ----------------------------------------------------------------------------------
52BH62_47...        ----------------------------------------------------------------------------------

                    1559                                                                                    1640
F10.seq             GCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGCAGCAAGGGGAGGAGTTGGGAGACA
52BH20_46...        ----------------------------------------------------------------------------------
52BH21_46...        ----------------------------------------------------------------------------------
52BH19_46...        ----------------------------------------------------------------------------------
43EH26_(4...        ----------------------------------------------------------------------------------
52BH07_46...        ----------------------------------------------------------------------------------
52BH08_46...        ----------------------------------------------------------------------------------
43EH26_(4...        ----------------------------------------------------------------------------------
52BH07_46...        ----------------------------------------------------------------------------AGACA
52BH63_47...        ----------------------------------------------------------------------------------
52BH61_47...        ----------------------------------------------------------------------------------
19H198_04...        ----------------------------------------------------------------------------------
52BH62_47...        ----------------------------------------------------------------------------------
```

```
                    1641                                                                                    1722
F10.seq             ATAGCAGGCCATGCTGGGGATCGCGTGGGCTCTATGGACTAGTctagtcaacaccggcccccttgatcttgtccactggccc
52BH20_46...        ----------------------------------------------------------------------------------
52BH21_46...        ----------------------------------------------------------------------------------
52BH19_46...        ----------------------------------------------------------------------------------
43EH26_(4...        ATAGCAGGCCATGCTGGGGATGCCGGTGGCTCTATGGACTAGTCTAGTCAACACCGGCCCCCTTGATCTTGTCCACTGGCCC
52BH07_46...        ----------------------------------------------------------------------------------
52BH08_46...        ----------------------------------------------------------------------------------
52BH63_47...        ----------------------------------------------------------------------------------
52BH61_47...        ----------------------------------------------------------------------------------
19H198_04...        ----------------------------------------------------------------------------------
52BH62_47...        ----------------------------------------------------------------------------------

                    1723                                                                                    1804
F10.seq             tggatcttgctggcaaagtcgcacagcgttgcggcagccactcttgagcaggtcagaaacactgcacgttggcaagccctt
52BH20_46...        ----------------------------------------------------------------------------------
52BH21_46...        ----------------------------------------------------------------------------------
52BH19_46...        ----------------------------------------------------------------------------------
43EH26_(4...        TGGATCTTGCTGGCAAAGTCGCACAGCGTTGCGGCAGCCACTCTTGAGCAGGTCAGAAACACTGCACGTTGGCAAGCCCTT
52BH07_46...        ----CAAAGGTCGCACAGCGTTGCGGCAGCCACTTCTTGAGCAGGTCAGAAACACTGCACGTTGGCAAGCCCTT
52BH08_46...        ----------------------------------------------------------------------------------
52BH63_47...        ----------------------------------------------------------------------------------
52BH61_47...        ----------------------------------------------------------------------------------
19H198_04...        ----------------------------------------------------------------------------------
52BH62_47...        ----------------------------------------------------------------------------------

                    1805                                                                                    1886
F10.seq             tgaggcagccagttgtgcagtccacacacagatcgacctgtgcgatgaactgtccatgggctccaagtccttgaacccagg
52BH20_46...        ----------------------------------------------------------------------------------
52BH21_46...        ----------------------------------------------------------------------------------
52BH19_46...        ----------------------------------------------------------------------------------
43EH26_(4...        TGAGGCAGCCAGTTGTGCAGTCCACACACAGATCGACCTGTGCGATGAACTGTCCATGGGCTCCAAGTCCTTGAACCCAGG
52BH07_46...        TGAGGCAGCCAGTTGTGCAGTCCACACACAGATCGACCTGTGCGATGAACTGTCTCATGGGCTCCAAGTCCTTGAACCCAGG
52BH08_46...        ----------------------------------------------------------------------------------
52BH63_47...        ----------------------------------------------------------------------------------
52BH61_47...        ----------------------------------------------------------------------------------
19H198_04...        ----------------------------------------------------------------------------------
52BH62_47...        ----------------------------------------------------------------------------------

                    1887                                                                                    1968
F10.seq             aatctcaggaatgtcgacgatcgcctcgcctgtgccggactcttgtcgccttcgtaagtgtggcagggtcct
52BH20_46...        ----------------------------------------------------------------------------------
52BH21_46...        ----------------------------------------------------------------------------------
52BH19_46...        ----------------------------------------------------------------------------------
43EH26_(4...        AATCTCAGGAATGTCGACGATCGCCTCGCCTATGCCGCCCGCTGTGCCGGACTCTTTGTCGCCTTCGTAGGTGTGGCAGGGTCT
52BH07_46...        AATCTCAGGAATGTCGACGATCGCCTCGCCTATGCCGCCCTGTGCCGGACTCTTTGTCGCCTTCGTAGGTGTGGCAGGGTCT
52BH08_46...        ----------------------------------------------------------------------------------
52BH63_47...        ----------------------------------------------------------------------------------
52BH61_47...        ----------------------------------------------------------------------------------
19H198_04...        ----------------------------------------------------------------------------------
52BH62_47...        ----------------------------------------------------------------------------------
```

Section 6.4

Page 7 52BH20_46967022_46967022 (52BH20_46967022_46967022.ab1), 52BH21_46967039_46967039 (52...

```
                    1969                                                                                                2050
F10_seq     gggatgaacttcattcattcttgggcgtgcaacttgatgtgggacaggcaagtcagacaagcgccctggtcagccagcgttccggg
52BH20_46..  ------------------------------------------------------------------------------------------
52BH21_46..  ------------------------------------------------------------------------------------------
52BH19_46..  ------------------------------------------------------------------------------------------
43EH26_(4..  GGGATGAACTTCTTCATTCTTGGGCGTGCACTTGATGTGTGGGACAGGCAGGCAGATCAGACAGCGCCCTGGTCAGCCAGCGTTCCGGG
52BH07_46..  GGGATGAACACTTCTTCATTCTTGGGCGTGCACTTGATGTGGGACAGGCAGGCAGATCAGACAGCGCCCTGGTCAGCCAGCGTTCCGGG
52BH08_46..  ------------------------------------------------------------------------------CTTTCCGGG
52BH63_47..  ------------------------------------------------------------------------------------------
52BH61_47..  ------------------------------------------------------------------------------------------
19H198_04..  ------------------------------------------------------------------------------------------
52BH62_47..  ------------------------------------------------------------------------------------------
..........

                    2051                                                                                                2132
F10_seq     cattggcttccatctcttgagcacctccagcgggcaagcttcttgccgggcaactccccgcggtcagcatcgaagatccgtggt
52BH20_46..  ------------------------------------------------------------------------------------------
52BH21_46..  ------------------------------------------------------------------------------------------
52BH19_46..  ------------------------------------------------------------------------------------------
43EH26_(4..  CATTGGCTTCCATCTCTTTGAGCACCTCCAGCGGGCAACTTCCCGCGGGTCAGCATCGAAGATCCGTGGT
52BH07_46..  CATTGGCTTCCATCTCTTTGAGCACCTCCAGCGGGCAACTTCCCGCGGGTCAGCATCGAAGATCCGTGGT
52BH08_46..  CATTGGCTTCCATCTCTTTGAGCACCTCCAGCGCGCAACTTCCCGCGGGTCAGCATCGAAGATCCGTGGT
52BH63_47..  ------------------------------------------------------------------------------------------
52BH61_47..  ------------------------------------------------------------------------------------------
19H198_04..  ------------------------------------------------------------------------------------------
52BH62_47..  ------------------------------------------------------------------------------------------
..........

                    2133                                                                                                2214
F10_seq     cgcgaagttgctggccacgccaacgatgttgaagtcttcgttgttctcggtgggttcatgccacgcacttccatggaggc
52BH20_46..  ------------------------------------------------------------------------------------------
52BH21_46..  ------------------------------------------------------------------------------------------
52BH19_46..  ------------------------------------------------------------------------------------------
43EH26_(4..  CGCGAAGTTGCTGGCCACGCCAACGATGTTGAAGTCTTCGTTGTTCTCGGTGGGTTCATGCCACGCACTTCCATGGAGGCC
52BH07_46..  CGCGAAGTTGCTGGCCACGCCAACGATGTTGAAGTCTTCGTTGTTCTCGGTGGGTTCATGCCACGCACTTCCATGGAGGCC
52BH08_46..  CGCGAAGTTGCTGGCCACGCCAACGATGTTGAAGTCTTCGTTGTTCTCGGTGGGTTCATGCCACGCACTTCCATGGAGGCC
52BH63_47..  ------------------------------------------------------------------------------------------
52BH61_47..  ------------------------------------------------------------------------------------------
19H198_04..  ------------------------------------------------------------------------------------------
52BH62_47..  ------------------------------------------------------------------------------------------
..........

                    2215                                                                                                2296
F10_seq     tgaagcagcaaggccagcaacaggccccgagccggctgcagcacagatctcggcatggtggttaattaagatctgtaacggc
52BH20_46..  ------------------------------------------------------------------------------------------
52BH21_46..  ------------------------------------------------------------------------------------------
52BH19_46..  ------------------------------------------------------------------------------------------
43EH26_(4..  TGAACCAGCAAGGCCAGCAACAGGCCCCGAGCCGGCTGCAGCACAGATCTCGGCATGGTGGTTAATTAAGATCTGTAACGGC
52BH07_46..  TGAACCAGCAAGGCCAGCAACAGGGCCCCGAGCCGGCTGCAGCACAGATCTCGGCATGGTGGTTAATTAAGATCTGTAACGGC
52BH08_46..  TGAACCAGCAAGGCCAGCAACAGGGCCCCGAGCCGGCTGCAGCACAGATCTCGGCATGGTGGTTAATTAAGATCTGTAACGGC
52BH63_47..  ------------------------------------------------------------------------------------------
52BH61_47..  ------------------------------------------------------------------------------------------
19H198_04..  ------------------------------------------------------------------------------------------
52BH62_47..  ------------------------------------------------------------------------------------------
..........
```

Page 8 52BH20_46967022_46967022 (52BH20_46967022_46967022.ab1), 52BH21_46967039_46967039 (52...

```
                    2297                                                                                                2378
F10_seq     GCAGAACAGAAAACGAAACCAAAAGACGTAGAGTTGAGCAAGCAGGGTCAGGCAAAGCGTGGAGAAGCCGGCTGAGTCTAGGTAG
52BH20_46..  ------------------------------------------------------------------------------------------
52BH21_46..  ------------------------------------------------------------------------------------------
52BH19_46..  ------------------------------------------------------------------------------------------
43EH26_(4..  GCAGAACAGAAAACGAAACCAAAAGACGTAGAGTTGAGCAAGCAGGGTCAGGCAAAGCGTGGAGAAGCCGGCTGAGTCTAGGTAG
52BH07_46..  GCAGAACAGAAAACGAAACCAAAAGACGTAGAGTTGAGCAAGCAGGGTCAGGCAAAGCGTGGAGAAGCCGGCTGAGTCTAGGTAG
52BH08_46..  GCAGAACAGAAAACGAAACCAAAAGACGTAGAGTTGAGCAAGCAGGGTCAGGCAAAGCGTGGAGAAGCCGGCTGAGTCTAGGTAG
52BH63_47..  ------------------------------------------------------------------------------------------
52BH61_47..  ------------------------------------------------------------------------------------------
19H198_04..  ------------------------------------------------------------------------------------------
52BH62_47..  ------------------------------------------------------------------------------------------
..........

                    2379                                                                                                2460
F10_seq     GCTTCAAGGGAGCGCCGGACAAAGGCCCGGTCTCGACCTGAGCTTTAAACTTACCTAGACGGGCGGACCGCAGTTCAGGAGGCA
52BH20_46..  ------------------------------------------------------------------------------------------
52BH21_46..  ------------------------------------------------------------------------------------------
52BH19_46..  ------------------------------------------------------------------------------------------
43EH26_(4..  GCTTCAAGGGAGCGCCGGACAAAGGCCCGGTCTCGACCTGAGCTTTAAACTTACCTAGACGGGCGGACCGCAGTTCAGGAGGCA
52BH07_46..  GCTTCAAGGGAGCGCCGGACAAAGGCCCGGTCTCGACCTGAGCTTTAAACTTACCTAGACGGGCG------------
52BH08_46..  GCTTCAAGGGAGCGCCGGACAAAGGCCCGGTCTCGACCTGAGCTTTAAACTTACCTAGACGGGCGGACCGCAGTTCAGGAGGCA
52BH63_47..  ------------------------------------------------------------------------------------------
52BH61_47..  ------------------------------------------------------------------------------------------
19H198_04..  ------------------------------------------------------------------------------------------
52BH62_47..  ------------------------------------------------------------------------------------------
..........

                    2461                                                                                                2542
F10_seq     CCACAGGCGGGAGGCGGCAGAACGCGACTCAAACCGGCGTGCGATGGCGGCCTCAGGTAGGCGCGGCGGGCGGCCGTGAAGGAGAGA
52BH20_46..  ------------------------------------------------------------------------------------------
52BH21_46..  ------------------------------------------------------------------------------------------
52BH19_46..  ------------------------------------------------------------------------------------------
43EH26_(4..  CCACAGGCGGGAGGCGGCAGAACGCGACTCAAACCGGCGTGCGATGGCGGCCTCAGGTAGGCGCGGCGGGCGGCCGTGAAGGAGAGA
52BH07_46..  CCACAGGCGGGAGGCGGCAGAACGCGACTCAAACCGGCGTGGATGGCGGCCTCAGGTAGGCGCGGCGGGCGGCCGTGAAGGAGAGA
52BH08_46..  CCACAGGCGGGAGGCGGCAGAACGCGACTCAAACCGGCGTGGATGGCGGCCTCAGGTAGGCGCGGCGGGCGGCCGTGAAGGAGAGA
52BH63_47..  ------------------------------------------------------------------------------------------
52BH61_47..  ------------------------------------------------------------------------------------------
19H198_04..  ------------------------------------------------------------------------------------------
52BH62_47..  ------------------------------------------------------------------------------------------
..........

                    2543                                                                                                2624
F10_seq     TGCCGACCCCTCGAAGCTTCAGCTGTGTTCTCGGCGGCAAACCCGTTCGCGAAAAAGAACGTTCACGGCGACTACTGCACTTAT
52BH20_46..  ------------------------------------------------------------------------------------------
52BH21_46..  ------------------------------------------------------------------------------------------
52BH19_46..  ------------------------------------------------------------------------------------------
43EH26_(4..  TGCCGACCCCTCGAAGCTTCAGCTGTGTTCTCGGCGGCAAACCCGTTCGCGAAAAAGAACGTTCACGGCGACTACTGCACTTAT
52BH07_46..  TGCCGACCCCTCGAAGCTTCAGCTGTGTTCTCGGCGGCAAACCCGTTCGCGAAAAAGAACGTTCACGGCGACTACTGCACTTAT
52BH08_46..  TGCCGACCCCTCGAAGCTTCAGCTGTGTTCTCGGCGGCAAACCCGTTCGCGAAAAAGAACGTTCACGGCGACTACTGCACTTAT
52BH63_47..  ------------------------------------------------------------------------------TTAT
52BH61_47..  ------------------------------------------------------------------------------------------
19H198_04..  ------------------------------------------------------------------------------------------
52BH62_47..  ------------------------------------------------------------------------------------------
..........
```

Page 9 52BH20_46967022_46967022 (52BH20_46967022_46967022.ab1), 52BH21_46967039_46967039 (52...

```
                  2625                                                                                          2706
F10_seq      ATACGGTTCTCCCCACCCTCGGGAAAAAGGGGGAGCCAGTACACGACATCACTTTCCCAGTTACCCGCGCCACTTCTC
52BH20_46... ATACGGTTCTCCCCACCCTCGGGAAAAAGGGGGAGCCAGTACACGACATCACTTTCCCAGTTACCCGCGCCACTTCTC ------
52BH21_46... |--------------------------------------------------------------------------------
52BH19_46... |--------------------------------------------------------------------------------
43EH26 (4.. ATACGGTTCTCCCCACCCTCGGGAAAAAGGGGGAGCCAGTACACGACATCACTTTCCCAGTTACCCGCGCCACTTCTC
52BH07_46... |--------------------------------------------------------------------------------
52BH08_46... ATACGGTTCTCCCCACCCTCGGGAAAAAGGGGGAGCCAGTACACGACATCACTTTCCCAGTTACCCGCGCCACTTCTC
52BH63_47... ATACGGTTCTCCCCACCCTGGGGAAAAAGGGGGAGCCAGTACACGACATCACTTTCCCAGTTACCCGCGCCACTTCTC
52BH61_47... ------AAAAGGCGGGAGCCAGTACACGACACTCACTTTCCCAGTTACCCGCGCCACTTCTC
19H198_04... |--------------------------------------------------------------------------------
52BH62_47... |--------------------------------------------------------------------------------
.............

                  2707                                                                                          2788
F10_seq      TAGGCACCGGTTCAATTGCCGACCCCTCCCCCAACTTCTCGGGGACTGTGGGCGGCATGTCGCGTCTGCCCACTGACGTCGAC
52BH20_46... TAGGCACCGGTTCAATTGCCGACCCCTCCCCCAACTTCTCGGGGACTGTGGGCGGCATGTCGCGTCTGCCCACTGACGTCGAC
52BH21_46... |--------------------------------------------------------------------------------------
52BH19_46... |--------------------------------------------------------------------------------------
43EH26 (4.. TAGGCACCGGTTCAATTGCCGACCCCTCCCCCAACTTCTCGGGGACTGTGGG---
52BH07_46... |--------------------------------------------------------------------------------------
52BH08_46... TAGGCACCGGTTCAATTGCCGACCCCTCCCCCAACTTCTCGGGGACTGTGGGCGGCATGTCGCGTCTGCCCACTGACGTCGAC
52BH63_47... TAGGCACCGGTTCAATTGCCGACCCCTCCCCCAACTTCTCGGGGACTGTGGGCGGCATGTCGCGTCTGCCCACTGACGTCGAC
52BH61_47... TAGGCACCGGTTCAATTGCCGACCCCTCCCCCAACTTCTCGGGGACTGTGGGCGGCATGTCGCGTCTGCCCACTGACGTCGAC
19H198_04... |--------------------------------------------------------------------------------------
52BH62_47... |--------------------------------------------------------------------------------------
.............

                  2789                                                                                          2870
F10_seq      AGTACTAAGCTTTGACAGAAAAGCCCCATCCTTAGGCCTCCTCCTCTTCCTAGTCTCCTGAATATTGGGTCTAACCCCACCTCC
52BH20_46... AGTACTAAGCTTTGACAGAAAAGCCCCATCCTTAGGCCTCCTCCTCTTCCTAGTCTCCTGAATATTGGGTCTAACCCCACCTCC
52BH21_46... |----------------------------------------------------------------------------------------
52BH19_46... |----------------------------------------------------------------------------------------
43EH26 (4.. |----------------------------------------------------------------------------------------
52BH07_46... |----------------------------------------------------------------------------------------
52BH08_46... AGTACTAAGCTTTGACAGAAAAGCCCCATCCTTAGGCCTCCTCCTCTTCCTAGTCTCCTGATATTGGGTCTAACCCCACCTCC
52BH63_47... AGTACTAAGCTTTGACAGAAAAGCCCCATCCTTAGGCCTCCTCCTCTTCCTAGTCTCCTGATATTGGGTCTAACCCCACCTCC
52BH61_47... AGTACTAAGCTTTGACAGAAAAGCCCCATCCTTAGGCCTCCTCCTCTTCCTAGTCTCCTGATATTGGGTCTAACCCCACCTCC
19H198_04... |----------------------------------------------------------------------------------------
52BH62_47... |----------------------------------------------------------------------------------------
.............

                  2871                                                                                          2952
F10_seq      TGTTAGGCAGATTCCTTATCTGGTGACACACCCCATTTCCTGGAGCCATCCTCTCTCCTTGCCAGAACCTCTAAGGTTTGCT
52BH20_46... TGTTAGGCAGATTCCTTATCTGGTGACACACCCCATTTCCTGGAGCCATCCTCTCTCCTTGCCAGAACCTCTAAGGTTTGCT
52BH21_46... |--------------------------------------------------------------------------------------
52BH19_46... |--------------------------------------------------------------------------------------
43EH26 (4.. |--------------------------------------------------------------------------------------
52BH07_46... |--------------------------------------------------------------------------------------
52BH08_46... |--------------------------------------------------------------------------------------
52BH63_47... TGTTAGGCAGATTCCTTATCTGGTGACACACCCCATTTCCTGGAGCCATCCTCTCTCCTTGCCAGAACCTCTAAGGTTTGCT
52BH61_47... TGTTAGGCAGATTCCTTATCTGGTGACACACCCCATTTCCTGGAGCCATCCTCTCTCCTTGCCAGAACCTCTAAGGTTTGCT
19H198_04... |--------------------------------------------------------------------------------------
52BH62_47... |--------------------------------------------------------------------------------------
.............
```

Page 10 52BH20_46967022_46967022 (52BH20_46967022_46967022.ab1), 52BH21_46967039_46967039 (5...

```
                  2953                                                                                          3034
F10_seq      TACGATGGAGCCAGAGAGGATCCTGGGAGGGAGAGAGCTTGGCAGGGGGTGGGAGGGGAAGAGGGGGGATGCGTGACCTGCCCGGT
52BH20_46... TACGATGGAGCCAGAGAGGATCCTGGGAGGGAGAGAGCTTGGCAGGGGGTGGGAGGGGAAGAGGGGGGATGCGTGACCTGCCCGGT
52BH21_46... |---------------------------------------------------------------------------------------------
52BH19_46... |---------------------------------------------------------------------------------------------
43EH26 (4.. |---------------------------------------------------------------------------------------------
52BH07_46... |---------------------------------------------------------------------------------------------
52BH08_46... |---------------------------------------------------------------------------------------------
52BH63_47... TACGGATGGAGCCAGAGAGGATCCTGGGAGGGAGAGAGCTTGGCAGGGGGTGGGAGGGGAAGAGGGGGGATCCGTGACCTGCCCCGGT
52BH61_47... TACGGATGGAGCCAGAGAGGATCCTGGGAGGGAGAGAGCTTGGCAGGGGGTGGGAGGGGAAGAGGGGGGATCCGTGACCTGCCCCGGT
19H198_04... |---------------------------------------------------------------------------------------------
52BH62_47... |---------------------------------------------------------------------------------------------
.............

                  3035                                                                                          3116
F10_seq      TCTCAGTGGCCACCCTGCGGCTACCCTCTCCCAGAACCTCGAGCTGCTCTGACGCGCGGGCGCGTCTGGTGCGTTTCACTGATCCTGG
52BH20_46... TCTCAGTGGCCACCCTGCGGCTACCCTCTCCCAGAACCTCGAGCTGCTCTGACGCGCGGGCGCGTCTGGTGCGTTTCACTGATCCTGG
52BH21_46... |-----------------------------------------------------------------------------------------------
52BH19_46... |-----------------------------------------------------------------------------------------------
43EH26 (4.. |-----------------------------------------------------------------------------------------------
52BH07_46... |-----------------------------------------------------------------------------------------------
52BH08_46... |-----------------------------------------------------------------------------------------------
52BH63_47... TCTCAGTGGCCACCCTGCGGCTACCCTCTCCCAGAACCTCGAGCTGCTCTGACGCGCGGGCGCGTCTGGTGCGTTTCACTGATCCTGG
52BH61_47... TCTCAGTGGCCACCCTGCGGCTACCCTCTCCCAGAATTTATATGTTCCTCCGTGGGTCAGTTTTTACCTGTGAGATAAGGCCAGTAGCCAG
19H198_04... |-----------------------------------------------------------------------------------------------
52BH62_47... |-----------------------------------------------------------------------------------------------
.............

                  3117                                                                                          3198
F10_seq      TGCTGCAGCTTCCTTACACTTCCCAGAGAGGAGGAAGCAGTTTGGAAAAAACAAAAATCAGAATAAGTTGGTCCTGAGTTCTAACT
52BH20_46... TGCTGCAGCTTCCTTACACTTCCCAGAGAGGAGGAAGCAGTTTGGAAAAAACAAAAATCAGAATAAGTTGGTCCTGAGTTCTAACT
52BH21_46... |------------------------------------------------------------------------------------------
52BH19_46... |------------------------------------------------------------------------------------------
43EH26 (4.. |------------------------------------------------------------------------------------------
52BH07_46... |------------------------------------------------------------------------------------------
52BH08_46... |------------------------------------------------------------------------------------------
52BH63_47... TGCTGCAGCTTCCTTACACTTCCCAGAGAGGAGGAAGCAGTTTGGAAAAAACAAAAATCAGAATAAGTTGGTCCTGAGTTCTAACT
52BH61_47... TGCTGCAGCCTTCCTTACACTTCCCAGAGAGGAGGAAGCAGTTTGGAAAAAACAAAAATCAGAATAAGTTGGTCCTGAGTTCTAACT
19H198_04... |------------------------------------------------------------------------------------------
52BH62_47... |------------------------------------------------------------------------------------------
.............

                  3199                                                                                          3280
F10_seq      TTGGCTCTTCACCTTTCTAGTCCCCAATTTATATATTGTTCCTCCGTGGGTCAGTTTTTACCTGTGAGATAAGGCCAGTAGCCAG
52BH20_46... TTGGCTCTTCACCTTTCTAGTCCCCAATTTATATATTGTTCCTCCGTGGGTCAGTTTTTACCTGTGAGATAAGGCCAGTAGCCAG
52BH21_46... |----------------------------------------------------------------------------------------
52BH19_46... |----------------------------------------------------------------------------------------
43EH26 (4.. |----------------------------------------------------------------------------------------
52BH07_46... |----------------------------------------------------------------------------------------
52BH08_46... |----------------------------------------------------------------------------------------
52BH63_47... TTGGCTCTTCACCTTTCTAGTCCCCAATTTATATATTGTTCCTCCGTGGGTCAGTTTTTACCTGTGAGATAAGGCCAGTAGCCAG
52BH61_47... TTGGCTCTTCACCTTTCTAGTCCCCAATTTATATATTGTTCCTCCGTGGGTCAGTTTTTACCTGTGAGATAAGGCCAGTAGCCAG
19H198_04... |----------------------------------------------------------------------------------------
52BH62_47... |----------------------------------------------------------------------------------------
.............
```

Page 11 52BH20_46967022_46967022 (52BH20_46967022_46967022.ab1), 52BH21_46967039_46967039 (5...

Page 12 52BH20_46967022_46967022 (52BH20_46967022_46967022.ab1), 52BH21_46967039_46967039 (5...

Figure 121: Sanger sequencing alignments of integrated donor DNA in IMR90-4 hPOMC-GLuc clone 6. Any mismatched bases are indicated in red (12 pages).

233

Section 6.4

6.4.11 IMR90-4 hPOMC-GLuc clone 4

Sanger sequencing alignment for IMR90-4 hPOMC-GLuc (4) donor DNA insertion. Traces corresponding to dark red sequencing arrows depicted in Figure 122. 52BH16: sequencing range #1, 52BH17: sequencing range #2, 52BH18: sequencing range #3, 43EH25: sequencing range #4, 19HI99: sequencing range #5, 52BH10: sequencing range #6, 52BH11: sequencing range #7, 52BH60: sequencing range #8 (10 pages).
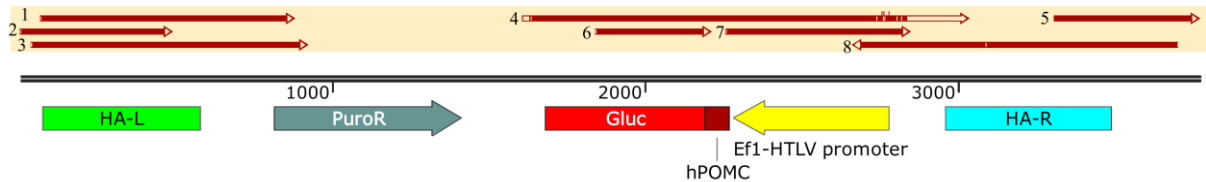
Figure 122: Sanger sequencing coverage of hPOMC-GLuc donor DNA inserted into AAVS1 safe harbor locus of IMR90-4 clone 4.

```
                 1                                                                                              82
C12_seq          TTCAGTTCCGTCTTCTCCACTCCTTGCTCTGCTGTGTGGTGTGCTGCCAAGGATGCTCTTTCCCGGAGCACTT
52BH17_46..      TTCAGTTCCGTCTTCTCCACTCCTTGCTCTGCTGTGTGGTGTGCTGCCAAGGATGCTCTTTCCCGGAGCACTT
52BH18_46..      ------TCCCTTCCCTTGCCTTGCTCTGCTGTGTGGTGTGCTGCCAAGGATGCTCTTTCCCGGAGCACTT
52BH16_46..      ------------TCCCTTGCCTTGCTCTGCTGTGTGGTGTGCTGCCAAGGATGCTCTTTCCCGGAGCACTT
52BH10_46..      ------------------GCTGTGTTGCTGCCAAGGATGCTCTTTCCCGGAGCACTT
43EH25_(4..      ---------------------------------------------------
52BH11_46..      ---------------------------------------------------
52BH60_47..      ---------------------------------------------------
19H199_04..      ---------------------------------------------------
```

```
                 83                                                                                             164
C12_seq          CCTTCTCCGGCGCCTGCACCACGTGATGTCCTCTGAGCGGATCCTCCCGTGTCTGGGTCTTCTCCGGGCATCTCCTCCCCCTC
52BH17_46..      CCTTCTCCGGCGCCTGCACCACGTGATGTCCTCTGAGCGGATCCTCCCGTGTCTGGGTCTTCTCCGGGCATCTCCTCCCCCTC
52BH18_46..      CCTTCTCCGGCCGCTGCACCACGTGATGTCCTCTGAGCGGATCCTCCCGTGTCTGGGTCTTCTCCGGGCATCTCCTCCCCCTC
52BH16_46..      CCTTCTCGGCGCCTGCACCACGTGATGTCCTCTGAGCGGATCCTCCCGTGTCTGGGTCATCTCCTCCCCCTC
52BH10_46..      CCTTCTCGGCGCTGCACCACGTGATGTCCTCTGAGCGGATCCTCCCGTGTCTGGGTCTCTCCGGGCATCTCCTCCCCCTC
43EH25_(4..      ---------------------------------------------------
52BH11_46..      ---------------------------------------------------
52BH60_47..      ---------------------------------------------------
19H199_04..      ---------------------------------------------------
```

```
                 165                                                                                            246
C12_seq          ACCCAACCCCATGCCCTCTTCACTCGCTGGGGTCCCCTTTTCCTTTCCTTCTCTCGGGCCATCTCTGTTTTCTTAGGA
52BH17_46..      ACCCAACCCCATGCCCTCTTCACTCGCTGGGGTCCCCTTTTCCTTTCCTTCTCTCGGGCCATCTCTGTTTTCTTAGGA
52BH18_46..      ACCCAACCCCATGCCGTCTTCACTCGCTGGGGTCCCCTTTTCCTTTCCTTCTCTCGGGCCATCTCTGTTTTCTTAGGA
52BH16_46..      ACCCACCCCATGCCGTCTTCACTCGCTGGGGTCCCCTTTTCCTTTCCTTCTCTCGGGCCATCTCTGTTTTCTTAGGA
52BH10_46..      ACCCACCCCATGCCGTCTTCACTCGCTGGGGTCCCCTTTTCCTTTCCTTCTCTCGGGCCATCTCTGTTTTCTTAGGA
43EH25_(4..      ---------------------------------------------------
52BH11_46..      ---------------------------------------------------
52BH60_47..      ---------------------------------------------------
19H199_04..      ---------------------------------------------------
```

```
                 247                                                                                            328
C12_seq          TGGCCTTCTCCGACGGAGTGTCTCCCGTGGGTCCGCGCTCCCCTTCTGTGTAGGCGCCATCATCACCGGTTTTCTGGACAACC
52BH17_46..      TGGCCTTCTCCGACGGAGTGTCTCCCGTGGGTCCGCGCTCCCCTTCTGTGTAGGCGCCATCATCACCGGTTTTCTGGACAACC
52BH18_46..      TGGCCTTCTCCGACGGAGTGTCTCCCGTGGGTCCGCGCTCCCCTTCTGTGTAGGCGCCATCATCACCGGTTTTCTGGACAACC
52BH16_46..      TGGCCTTCTCCGACGGAGTGTCTCCCGTGGGTCCGCGCTCCCCTTCTGTGTAGGCGCCATCATCACCGGTTTTCTGGACAACC
52BH10_46..      TGGCCTTCTCCGACGGATGTCTCCCGTGGGTCCGCGCTCCCCTTCTGTGTAGGCGCCATCATCACCGGTTTTCTGGACAACC
43EH25_(4..      ---------------------------------------------------
52BH11_46..      ---------------------------------------------------
52BH60_47..      ---------------------------------------------------
19H199_04..      ---------------------------------------------------
```

```
                 329                                                                                            410
C12_seq          CCAAAGTACCCCGTCTCCCTGGCTTTAGCCACCTCTCCATCCTCCTTGGTCTTTCTTTGCCTGGACACCCCGTTCTCCTGTGGAT
52BH17_46..      CCAAAGTACCCCGTCTCCCTGGCTTTAGCCACCTCTCCATCCTCCTTGGTCTTTCTTTGCCTGGACACCCCGTTCTCCTGTGGAT
52BH18_46..      CCAAAGTACCCCGTCTCCCTGGCTTTAGCCACCTCTCCATCCTCCTTGGTCTTTCTTTGCCTGGACACCCCGTTCTCCTGTGGAT
52BH16_46..      CCAAAGTACCCCGTCTCCCTGGCTTTAGCCACCTCTCCATCCTCCTTGGTCTTTCTTTGCCTGGACACCCCGTTCTCCTGTGGAT
52BH10_46..      CCAAAGTACCCCGTCTCCCTGGCTTTAGCCACCTCTCCATCCTCCTTGGTCTTTCTTTGCCTGGACACCCCGTTCTCCTGTGGAT
43EH25_(4..      ---------------------------------------------------
52BH11_46..      ---------------------------------------------------
52BH60_47..      ---------------------------------------------------
19H199_04..      ---------------------------------------------------
```

```
                 411                                                                                            492
C12_seq          TCGGGTCAACCTCTCACTCCTTTTCATTTGGGCAGCTCCCTACCCCCTTACCCTCTCTAGTCTGTGCTAGCTCTTCCAGCCCC
52BH17_46..      TCGGGTCAACCTCTCACTCCTTTTCATTTGGGCAGCTCCCTACCCCCTTACCCTCTCTAGTCTGTGCTAGCTCTTCCAGCCCC
52BH18_46..      TCGGGTCAACCTCTCACTCCTTTCATTTGGGCAGCTCCCTACCCCCTTACCCTCTCTAGTCTGTGCTAGCTCTTCCAGCCCC
52BH16_46..      TCGGGTCAACCTCTCACTCCTTTCATTTGGGCAGCTCCCTACCCCCTTACCCTCTCTAGTCTGTGCTAGCTCTTCCAGCCCC
52BH10_46..      TCGGGTCAACCTCTCACTCCTTTCATTTGGGCAGCTCCCTACCCCCTTACCCTCTCTAGTCTGTGCTAGCTCTTCCAGCCCC
43EH25_(4..      ---------------------------------------------------
52BH11_46..      ---------------------------------------------------
52BH60_47..      ---------------------------------------------------
19H199_04..      ---------------------------------------------------
```

```
                 493                                                                                            574
C12_seq          CTGTCATGGGCATCTTCCAGGGGTCCGAGAGCTCAGCTAGTCTTCTTCCTCCAACCCGGGGCCCCTATGTCACTCAGGACAG
52BH17_46..      CTGTCA------------------------------------------------------
52BH18_46..      CTGTCATGGGCATCTTCCAGGGGTCCGAGAGCTCAGCTAGTCTTCTTCCTCCAACCCGGGGCCCCTATGTCACTCAGGACAG
52BH16_46..      CTGTCATGGGCATCTTCCAGGGGTCCGAGAGCTCAGCTAGTCTTCTTCCTCCAACCCGGGGCCCCTATGTCACTCAGGACAG
52BH10_46..      CTGTCATGGGCATCTTCCAGGGGTCCGAGAGCTCAGCTAGTCTTCTTCCTCCAACCCGGGGCCCCTATGTCACTCAGGACAG
43EH25_(4..      ---------------------------------------------------
52BH11_46..      ---------------------------------------------------
52BH60_47..      ---------------------------------------------------
19H199_04..      ---------------------------------------------------
```

```
                 575                                                                                            656
C12_seq          CATGTTTGCTGCCTCCAGGATCCTGTCCCCGAGCTGGACACCTTATATTCCCAGGGGCGGGTTAAGTGGCTCTGGTT
52BH17_46..      CATGTTTGCTGCCTCCAGGATCCTGTCCCCGAGCTGGACACCTTATATTCCCAGGGGCGGGTTAAGTGGCTCTGGTT
52BH18_46..      CATGTTTGCTGCCTCCAGGATCCTGTCCCCGAGCTGGACACCTTATATTCCCAGGGGCGGGTTAAGTGGCTCTGGTT
52BH16_46..      CATGTTTGCTGCCTCCAGGATCCTGTCCCCGAGCTGGACACCTTATATTCCCAGGGGCGGGTTAAGTGGCTCTGGTT
52BH10_46..      CATGTTTGCTGCCTCCAGGATCCTGTCCCCGAGCTGGACACCTTATATTCCCAGGGGCGGGTTAAGTGGCTCTGGTT
43EH25_(4..      ---------------------------------------------------
52BH11_46..      ---------------------------------------------------
52BH60_47..      ---------------------------------------------------
19H199_04..      ---------------------------------------------------
```

```
                 657                                                                                            738
C12_seq          CTGGGTACTTTTATCTGTCCCCTCCACCCCACAGTGGGCGCAAGTCTTCTGACCTCTTCTCTCTCTCCCACAGGGCCTCGAGAG
52BH17_46..      CTGGGTACTTTTATCTGTCCCCTCCACCCCACAGTGGGCGCAAGTCTTCTGACCTCTTCTCTCTCTCCCACAGGGCCTCGAGAG
52BH18_46..      CTGGGTACTTTTATCTGTCCCCTCCACCCCACAGTGGGCGCAAGTCTTCTGACCTCTTCTCTCTCTCCCACAGGGCCTCGAGAG
52BH16_46..      CTGGGTACTTTTATCTGTCCCCTCCACCCCACAGTGGGCGCAAGTCTTCTGACCTCTTCTCTCTCTCCCACAGGGCCTCGAGAG
52BH10_46..      CTGGGTACTTTTATCTGTCCCCTCCACCCCACAGTGGGCGCAAGTCTTCTGACCTCTTCTCTCTCTCCCACAGGGCCTCGAGAG
43EH25_(4..      ---------------------------------------------------
52BH11_46..      ---------------------------------------------------
52BH60_47..      ---------------------------------------------------
19H199_04..      ---------------------------------------------------
```

```
                 739                                                                                            820
C12_seq          ATCTGGCAGCCGGAGGAGGGCAGAGGAAGTCTTCTAACATGCGGTGACGTGAGGAGGAGAATCCGGCCTCAGGCTCGAGATGACC
52BH17_46..      ATCTGGCAGCCGGAGGAGGGCAGAGGAAGTCTTCTAACATGCGGTGACGTGAGGAGGAGAATCCGGCCTCAGGCTCGAGATGACC
52BH18_46..      ATCTGGCAGCCGGAGGAGGGCAGAGGAAGTCTTCTAACATGCGGTGACGTGAGGAGGAGAATCCGGCCTCAGGCTCGAGATGACC
52BH16_46..      ATCTGGCAGCCGGAGGAGGGCAGAGGAAGTCTTCTAACATGCGGTGACGTGAGGAGGAGAATCCGGCCTCAGGCTCGAGATGACC
52BH10_46..      ATCTGGCAGCCGGAGGAGGGCAGAGGAAGTCTTCTAACATGCGGTGACGTGAGGAGGAGAATCCGGCCTCAGGCTCGAGATGACC
43EH25_(4..      ---------------------------------------------------
52BH11_46..      ---------------------------------------------------
52BH60_47..      ---------------------------------------------------
19H199_04..      ---------------------------------------------------
```

235

Section 6.4

```
                   821                                                                              902
C12_seq            GAGTACAAGCCACGGTGCGGCCTCGCCCACCCGACGACGAAGCTCCCCAGGCCGTACGCGACCCTCGCCGCGCCGGGTTCGCCGACT
52BH17_46...
52BH18_46...       GAGTACAAGCCCACCACGGTGCGGGCCTCGCCACCCGACGACGAAGCTCCCCAGGCCGTACGCGACCCTCGCCGCGCCGGGTTCGCCGACT
52BH16_46...
52BH10_46...       GAGTACAAGCCCACCACGGTGCGGGCCTCGCCACCCGACGACGAAGCTCCCCAGGCCGTACGCGACCCTCGCCGCGCCGGGTTCGCCGACT
43EH25_(4.
52BH25_(4.
52BH11_46.
52BH60_47.
19H199_04.

                   903                                                                              984
C12_seq            ACCCCGCCACCGCCACCACGCGTCGATCCGGACCGGCGTCACCGCGTCACCGAGCTGCAAGAACTCTTCCTCACGCGCGT
52BH17_46...
52BH18_46...       ACCCCGCCACCGCGACCACACCGTCG
52BH16_46...       ACCCCGCCACCGCCACACCGTCGATCCGGACCGGCGCCACACCGAGCTGCAAGAACTCTTCCTCACGCGCGT
52BH10_46...
43EH25_(4.
52BH25_(4.
52BH11_46.
52BH60_47.
19H199_04.

                   985                                                                              1066
C12_seq            CGGGCTCGACATCGGCAAGGIGTGGGTCGGGCGACGACGCGCCGCGGCGTGGACCACGCCGGAGAGCGCCGTCGAAGCG
52BH17_46...
52BH18_46...
52BH16_46...       CGGGCTCGACATCGGCAAGGTGTGGGTCGCGCGACGCGACGGCGCCGC
52BH10_46...
43EH25_(4.
52BH25_(4.
52BH11_46.
52BH60_47.
19H199_04.

                   1067                                                                             1148
C12_seq            GGGGCGGTGTTCGCCGAGATCGGCCCCGCGCCATGGCCGAGTTGAGCCGGTTCCCGCGGCTGGCCGCCGCAGCAACAGAATGGAAGGCC
52BH17_46...
52BH18_46...
52BH16_46...
52BH10_46...
43EH25_(4.
52BH25_(4.
52BH11_46.
52BH60_47.
19H199_04.

                   1149                                                                             1230
C12_seq            TCCTGGCCGCCGCACCGCGCCCAAGGAGCCCCCGTGGTTCCTGGCCACCGTCCGCCCGACCCACCAGGGGCAAGGGTCT
52BH17_46...
52BH18_46...
52BH16_46...
52BH10_46...
43EH25_(4.
52BH25_(4.
52BH11_46.
52BH60_47.
19H199_04.
```

```
                   1231                                                                             1312
C12_seq            GGGGCAGCGCCGTCGTGCTCCCCGGAGTGGAGGCGGCGCCGAGCGACGGCGGGGTGCCCGCCTTCCTGGAGACCTCCGCCGCCCGC
52BH17_46...
52BH18_46...
52BH16_46...
52BH10_46...
43EH25_(4.
52BH25_(4.
52BH11_46.
52BH60_47.
19H199_04.

                   1313                                                                             1394
C12_seq            AACCTCCCCTTCTACGAGCCGGGCTCGGCTTCACCGTCACCGCGCCGACGTCGAGGTGCCCGGAGACGGCGCACCTGGTCGCATGA
52BH17_46...
52BH18_46...
52BH16_46...
52BH10_46...
43EH25_(4.
52BH25_(4.
52BH11_46.
52BH60_47.
19H199_04.

                   1395                                                                             1476
C12_seq            CCCGCAAGCCCGGTGCCTGATCTAGACGGCCCGTTTAAAACCCGCTGATCAGCCTCGACTGTGCCTTCTTCTAGTTGCCAGCCATC
52BH17_46...
52BH18_46...
52BH16_46...
52BH10_46...
43EH25_(4.
52BH25_(4.
52BH11_46.
52BH60_47.
19H199_04.

                   1477                                                                             1558
C12_seq            TGTTGTTTGCCCTCCCCGCTCCTTGACCCTGGAAGGTGCCACTCCCACTCCTTTCCTTCCTAATAAAATGAGGAAATT
52BH17_46...
52BH18_46...
52BH16_46...
52BH10_46...
43EH25_(4.                                                                                           TCAGGAAATT
52BH25_(4.
52BH11_46.
52BH60_47.
19H199_04.

                   1559                                                                             1640
C12_seq            GCATCGCATTGTCTCAGTAGGTGTCATTCTATTCTGGGGGTGGGGCTGGGGCGCAGGAGACAGCACAGCAGGGGGAGGAGGATTGGGAAGACA
52BH17_46...
52BH18_46...
52BH16_46...
52BH10_46...       GCATCGCATTGTCTCAGTAGGTGTCATTCTATTCTGGGGGTGGGGCTGGGGCGCAGGAGACACAGCAGGGGGAGGAGGATTGGGAAGACA
43EH25_(4.
52BH25_(4.
52BH11_46.                                                                                          AGACA
52BH60_47.
19H199_04.
```

```
                1641                                                                                          1722
C12_seq         ATAGCAGGCATGCTGGGGATGCGGTGGGCGGTCTCTATCGACTAGTGTctagtcaccaccaccggccccccttgatcttgtccaactcggccc
52BH17_46..     ATAGCAGGCATGCTGGGGATGCGGTGGGCGGTCTCTATCGACTAGTGTctagtcaccaccaccggccccccttgatcttgtccaactcggccc
52BH18_46..     ---------------------------------------------------------------------------------------------
52BH16_46..     ---------------------------------------------------------------------------------------------
52BH10_46..     ATAGCAGGCATGCTGGGGATGCGGTGGGCGGTCTCTATCGACTAGTGTACCACCGGCCCCCCTTGATCTTGTCCAACTCGGTGGT
43EH25 (4..     ATAGCAGGCATGCTGGGGATGGCGGTGGGCGGTCTCTATCGACTAGTGTCACCACCGGCCCCCCTTGATCTTGTCCAACTCCGTGGT
52BH11_46..     ATAGCAGGCATGCTGGGGATGGCGGTGGGCGGTCTCTATCGACTAGTGTCACCACCGGCCCCCCTTGATCTTGTCCAACTCCGTGCC
52BH60_47..     ---------------------------------------------------------------------------------------------
19H199_04..     ---------------------------------------------------------------------------------------------
                                        . . . . . . . . . . . . . . . . . . . . . . . .

                1723                                                                                          1804
C12_seq         tggatcttgctggcaaaggtcgcacaacgttgcggcagccaactcttcttgaagcaaggtcagaacactgcaagttggcaagccctt
52BH17_46..     tggatcttgctggcaaaggtcgcacaacgttgcggcagccaactcttcttgaagcaaggtcagaacactgcaagttggcaagccctt
52BH18_46..     ---------------------------------------------------------------------------------------------
52BH16_46..     ---------------------------------------------------------------------------------------------
52BH10_46..     TGGATCTTGCTGGCAAAGGTCGCACAACGTTCCGGCAGCCAACTCTTCTTGAAGCAGGTCAGAACACTGCACGTTGGCAAGCCCTT
43EH25 (4..     TGGATCTTGCTGGCAAAGGTCGCACAACGTTGCCGGCAGCCACTCTTCTTGAAGCAGGTCAGAACACTGCACGTTGGCAAGCCCTT
52BH11_46..     ---------------------------------------------------------------------------------------------
52BH60_47..     ---------------------------------------------------------------------------------------------
19H199_04..     ---------------------------------------------------------------------------------------------
                                        . . . . . . . . . . . . . . . . . . . . . . . .

                1805                                                                                          1886
C12_seq         tgaggcagccagttgtcagtccacacacaagatcgaacctgtgcgatgaactgctccatggctccaagtcctttgaaccccagg
52BH17_46..     tgaggcagccagttgtcagtccacacacaagatcgaacctgtgcgatgaactgctccatggctccaagtcctttgaaccccagg
52BH18_46..     ---------------------------------------------------------------------------------------------
52BH16_46..     ---------------------------------------------------------------------------------------------
52BH10_46..     TGAGGCAGCCAGTTGTTCAGTCCACACACAAGATCGAACCTGTCCATCGGCTCCATGGGCTCCATGGGCTCCAAGTCCTTTGAACCCAGG
43EH25 (4..     TGAGGCAGCCAGTTGTTGCAGTCCACACACAGATCGAACCTGTGCGATGAACTGCTCCATGGGCTCCAAGTCCTTTGAACCCAGG
52BH11_46..     TGAAGCAGCCAAGGCCAGCAACAGGGCCCCCAGCGGCTGCAACACGAAGTCTTCCGTGTTCTCGGTGGGCTTAATTAAGATCTTGTAACGGC
52BH60_47..     ---------------------------------------------------------------------------------------------
19H199_04..     ---------------------------------------------------------------------------------------------
                                        . . . . . . . . . . . . . . . . . . . . . . . .

                1887                                                                                          1968
C12_seq         aatttcaggaatgtcgacgatcgcctcgcctatgccgcccctgtgcggaactcctttgtcgccttcgtaggtgttggaagcgtcct
52BH17_46..     aatttcaggaatgtcgacgatcgcctcgcctatgccgcccctgtgcggaactcctttgtcgccttcgtaggtgttggaagcgtcct
52BH18_46..     ---------------------------------------------------------------------------------------------
52BH16_46..     ---------------------------------------------------------------------------------------------
52BH10_46..     AATTCAGGAATGTCGACCATGCCCTCGCCTATGCCGCCCCTGTGCGGAACTCTTTGTCGCCTTCGTAGGTGTTGGCACGTCT
43EH25 (4..     AATCCAGGAATGTCGACGATCGCCTCGCCTATCCGCCCGTGTCGCGGAACTCTTTGTCGCCTTCGTAGGTGTTGGCACGTCTCT
52BH11_46..     ---------------------------------------------------------------------------------------------
52BH60_47..     ---------------------------------------------------------------------------------------------
19H199_04..     ---------------------------------------------------------------------------------------------
                                        . . . . . . . . . . . . . . . . . . . . . . . .

                1969                                                                                          2050
C12_seq         gggatgaacttcttcattcttgggcgtgcaacttgatgtgggacaggcaagatcagacagcagccctggtgcagccaagttccggg
52BH17_46..     gggatgaacttcttcattcttgggcgtgcaacttgatgtgggacaggcaagatcagacagcagccctggtgcagccaagttccggg
52BH18_46..     ---------------------------------------------------------------------------------------------
52BH16_46..     ---------------------------------------------------------------------------------------------
52BH10_46..     GGGATGAACTTCTTCATCTTGGGCGTGCACTTGATGTGGGACAGGCAGATCAGACAGCGCCCTGGTGCAGCCAGTTCCGGG
43EH25 (4..     GGGATGAACTTCTTCATCTTGGGCGTGCACTTGATGTGGGGACAGGCAGGATCAGACAGCAGCCCTGGTGCAGCCAGTTCAGGAGGCA
52BH11_46..     ---------------------------------------------------------------------------------------------
52BH60_47..     ---------------------------------------------------------------------------------------------
19H199_04..     ---------------------------------------------------------------------------------------------
                                        . . . . . . . . . . . . . . . . . . . . . . . .
```

```
                2051                                                                                          2132
C12_seq         cattggcttccatctccttgagcaacctccagcggcagcttcttgccgggcaacctcccgcggtcagcatcgagagatccgggt
52BH17_46..     cattggcttccatctccttgagcaacctccagcggcagcttcttgccgggcaacctcccgcggtcagcatcgagagatccgggt
52BH18_46..     ---------------------------------------------------------------------------------------------
52BH16_46..     ---------------------------------------------------------------------------------------------
52BH10_46..     CATTGGCTTCCATCTCTTTGAGCACCTCCAGCGCGGCCAGCTTCTTGCCGGGCAACTTCCCGCGGTCAGCATCGAGAGATCCGTGGT
43EH25 (4..     CATTGGCTTCCATCTCTTTGAGCACCTCCAGCGCGGCCAGCTTCTTGCCGGGCAACTTCCCGCGGTCAGCATCGAGAGATCCGTGGT
52BH11_46..     ----TTTGAGCACCTCCAGCGCGGCCAGCTTCTTGCCGGGCAACTTCCCGCGGTCAGCATCGAGAGATCCGTGGT
52BH60_47..     ---------------------------------------------------------------------------------------------
19H199_04..     ---------------------------------------------------------------------------------------------
                                        . . . . . . . . . . . . . . . . . . . . . . . .

                2133                                                                                          2214
C12_seq         cgcgaagttgctggcacgcacgcacgatgttgaagtcttcgttcgttctcggtgggcttcatcgcacgcacttccatggaggc
52BH17_46..     cgcgaagttgctggcacgcacgcacgatgttgaagtcttcgttcgttctcggtgggcttcatcgcacgcacttccatggaggc
52BH18_46..     ---------------------------------------------------------------------------------------------
52BH16_46..     ---------------------------------------------------------------------------------------------
52BH10_46..     CGCGAAGTTGCTGGCCACGGCCACGATGTTGAAGTCTTCGTTCGTTCTCGGTGGGCTTCATCGCCACGCACTTCCATGGAGGCC
43EH25 (4..     CGCGAAGTTGCTGGCCACGGCCACGATGTTGAAGTCTTCGTTCGTTCTCGGTGGGCTTCATCGCCACGCACTTCCATGGAGGCC
52BH11_46..     CGCGAAGTTGCTGGCCACGGCCACGATGTTGAAGTCTTCGTTGTTCTCGGTGGGCTTCATCGCCACGCACTTCCATGGAGGCC
52BH60_47..     ---------------------------------------------------------------------------------------------
19H199_04..     ---------------------------------------------------------------------------------------------
                                        . . . . . . . . . . . . . . . . . . . . . . . .

                2215                                                                                          2296
C12_seq         tgaagcagcaaggccagcacaagggccccgagcggctgcagcacagatctcggcacatggtggTTAATTAAGATCTGTAACGGC
52BH17_46..     tgaagcagcaaggccagcacaagggccccgagcggctgcagcacagatctcggcacatggtggTTAATTAAGATCTGTAACGGC
52BH18_46..     ---------------------------------------------------------------------------------------------
52BH16_46..     ---------------------------------------------------------------------------------------------
52BH10_46..     TGAAGCAGCAAGGCCAGCACAAGGGCCCCGAGCGGCTGCAGCACAGATCTCGGCACATGGTGGTTAATTAAGATCTGTAACGGC
43EH25 (4..     TGAAGCAGCAAGGCCAGCACAAGGGCCCCGAGCGGCTGCAGCACAGATCTCGGCACATGGTGGTTAATTAAGATCTGTAACGGC
52BH11_46..     TGAAGCAGCAAGGCCAGCACAAGGGCCCCGAGCGGCTGCAGCACAGATCTCGGCACATGGTGGTTAATTAAGATCTGTAACGGC
52BH60_47..     ---------------------------------------------------------------------------------------------
19H199_04..     ---------------------------------------------------------------------------------------------
                                        . . . . . . . . . . . . . . . . . . . . . . . .

                2297                                                                                          2378
C12_seq         GCAGAACAGAAAACGAAACAAAAGACGTAGAGTTGAGCAACCACAGGGTCAGGCGCAAACGTGGAGAGACCGGCTCAGTCTCAGGTAG
52BH17_46..     GCAGAACAGAAAACGAAACAAAAGACGTAGAGTTGAGCAACCACAGGGTCAGGCGCAAACGTGGAGAGACCGGCTCAGTCTCAGGTAG
52BH18_46..     ---------------------------------------------------------------------------------------------
52BH16_46..     ---------------------------------------------------------------------------------------------
52BH10_46..     GCAGAACAGAAAACGAAACAACGTAGAGTTGAGCAAGCAGGGTCAGGCGCAAACGTGGAGAGACCGGCTGAGTCTCAGGTAG
43EH25 (4..     GCAGAACAGAAAACGAAACAACGTAGAGTTGAGCAAGCAGGGTCAGGCGCAAACGTGGAGAGACCGGCTGAGTCTCAGGTAG
52BH11_46..     GCAGAACAGAAAACGAAACAACGTAGAGTTGAGCAAGCAGGGTCAGGCGCAAACGTGGAGAGACCGGCTGAGTCTCAGGTAG
52BH60_47..     ---------------------------------------------------------------------------------------------
19H199_04..     ---------------------------------------------------------------------------------------------
                                        . . . . . . . . . . . . . . . . . . . . . . . .

                2379                                                                                          2460
C12_seq         GCTCCAAGGGAGCGCCGGACAAAGGCCCGGTCTCGACCTCAGCTTTAAACTTACCTACGGCGGCGGCGACCGCAGTTCAGGAGGCA
52BH17_46..     GCTCCAAGGGAGCGCCGGACAAAGGCCCGGTCTCGACCTCAGCTTTAAACTTACCTACGGCGGCGGCGACCGCAGTTCAGGAGGCA
52BH18_46..     ---------------------------------------------------------------------------------------------
52BH16_46..     ---------------------------------------------------------------------------------------------
52BH10_46..     GCTCCAAGGGAGCGCCGGACAAAGGCCCGGTCTCGACCTCAGCTTTAAACTTACCTAGACGGGCGGACGGACCGCAGTTCAGGAGGCA
43EH25 (4..     GCTCCAAGGGAGCGCCGGACAAAGGCCCGGTCTCGACCTCAGCTTTAAACTTACCTAGACGGGCGGACGGACCGCAGTTCAGGAGGCA
52BH11_46..     GCTCCAAGGGAGCGCCGGACAAAGGCCCGGTCTCGACCTCAGCTTTAAACTTACCTAGACGGGCGGACGGACCGCAGTTCAGGAGGCA
52BH60_47..     ---------------------------------------------------------------------------------------------
19H199_04..     ---------------------------------------------------------------------------------------------
                                        . . . . . . . . . . . . . . . . . . . . . . . .
```

237

Section 6.4

```
                2461                                                                                              2542
C12_seq      CCACAGGCGGGAGGCGGCAGGAACGCGACTCAACCGCGCTGGATGGCGGCCTCAGGTAGGCGGCGGCGGCCGGTGAAGGAGAGA
52BH17_.46..
52BH18_.46..
52BH16_.46..
52BH16_.46..
52BH10_.46.. CCACAGGCGGGAGGCGGCAGGAACGCGACTCAACCGCGCTGGATGGCGGCCTCAGGTAGGCGGCGGCGGCCGGTGAAGAGAGAGA
43EH25_.(4.. CCACAGGCGGGAGGCGGCAGGAACGCGACTCAACCGCGCTGGATGGCGGCCTCAGGTAGGCGGCGGCGGCCGGTGAAGAGAGAGA
52BH11_.46.. CCACAGGCGGGAGGCGGCAGGAACGCGACTCAACCGCGCTGGATGGCGGCCTCAGGTAGGCGGCGGCGGCCGGTGAAGGAGAGA
52BH60_.47..
19H199_04..
```

```
                2543                                                                                              2624
C12_seq      TGCCAGCCCCTCGGAAGCTTCACTCTGTCTTCTCTGGCGGCGCAAAACCCGTTGCGGAAAAAGAACGTTCACGCGACTACTGCACTTAT
52BH17_.46..
52BH18_.46..
52BH16_.46..
52BH16_.46..
52BH10_.46.. TGCCAGCCCCTCGAAAGCTTCAGCTGTGTTCTCTGGCGGCGCAAAACCCGTTGCGGAAAAAAGAACGTTCACGCGACTACTGCACTTAT
43EH25_.(4.. TGCCAGCCCCTCGAAAGCTTCAGCTGTGTTCTCTGGCGGCGCAAAACCCGTTGCGGAAAAAAGAACGTTCACGCGACTACTGCACTTAT
52BH11_.46.. TGCCAGCCCCTCGAAGCTTCACTGTGTTCTCTGGCGGCGCAAAACCCGTTGCGGAAAAAGAACGTTCACGCGACTACTGCACTTAT
52BH60_.47.. -------------------------------------------------------------------------------GCACTTAT
19H199_04..  -------------------------------------------------------------------------------GCACTTAT
```

```
                2625                                                                                              2706
C12_seq      ATACGGTTCTCCCCCACCCTCGGGAAAAAGGCGGAGCCACTACACGACATCACTTCCCACGTTACCCGCGCCCACCTTCTC
52BH17_.46..
52BH18_.46..
52BH16_.46..
52BH16_.46..
52BH10_.46.. ATACGGTTCTCCCCCACCCTCGGGAAAAAAGGGGGAGCCAGTACACGACATCACTTTCCCACGTTACCCGGCGCCCACCTTCTC
43EH25_.(4.. ATACGGTTCTCCCCCACCCTCGGGAAAAAAGGGGGAGCCAGTACACGACATCACTTTCCCACGTTACCCGGCGCCCACCTTCTC
52BH11_.46.. ATACGGTTCTCCCCCACCTCGGGAAAAAGGGGGAGCCAGTACACGACATCACTTTCCCACGTTACCCGGCGCCCACCTTCTC
52BH60_.47.. ATACGGTCTCCCCCACCCTCGGGAAAAAGGGGGGAGCCAGTACACGACATCACTTTCCCACGTTACCCGCGCCCACCTTCTC
19H199_04..
```

```
                2707                                                                                              2788
C12_seq      TAGGCACCGGTTCAATTGCCGACCCTCCCCCCAACTTCTCGGGGACTGTGGGCGATGCGCGCTCTGCCCACTGACCTGCGAC
52BH17_.46..
52BH18_.46..
52BH16_.46..
52BH16_.46..
52BH10_.46.. TAGGCACCGGTTCAATTGCCGACCCTCCCCCCCAACTTCTCGGGGACTGTGGG------
43EH25_.(4.. TAGGCACCGGTTCAATTGCCGACCCTCCCCCCAACTTCTCGGGGACTGTGGGCGATGTGCGCTCTGCCCACTGACGTGAC
52BH11_.46.. TAGGCACCGGTTCAATTGCCGACCCCTCCCCCCAACTTCTCGGGGACTGTTGGGCGATAGTGCGCTCTGCCCACTGACGTGAC
52BH60_.47.. TAGGCACCGGTTCAATTGCCGACCCCTCCCCCCAACTTCTCGGGGACTGTTGGGCGATGTGCGCTCTGCCCACTGACGTGAC
19H199_04..
```

```
                2789                                                                                              2870
C12_seq      AGTACTAAGCTTTGACAGAAAAGCCCATCCTTAGGCCTCCTCCTTCCTATCTCCTCGATATTGGGTCTAACCCCACCTCC
52BH17_.46..
52BH18_.46..
52BH16_.46..
52BH16_.46..
52BH10_.46..
43EH25_.(4.. AGTACTAAGCTTTGACAGAAAAGCCCATCCTTAGGCCTCCTCCTTCCTAGTCTCCTCGATATTGGGTCTAACCCC----
52BH11_.46.. AGTACTAAGCTTTGACAGAAAAGCCCATCCTTAGGCCTCCTCCTTCCTAGTCTCCTCGATATTGGGTCTAACCCCC----
52BH10_.46.. AGTACTAAGCTTTGACAGAAAAGCCCATCCTTAGGCCTCCTCCTTCCTAGTCTCCTCGATATTGGGTCTAACCCCC----
52BH60_.47.. AGTACTAAGCTTGACAGAAAAGCCCCATCCTTAGGCCCTCCTCCTTCCTAGTCTCCTCGATATTGGGTCTAACCCCCACCTCC
19H199_04..  AGTACTAAGCTTGACAGAAAAGCCCCATCCTTAGGCCCTCCTCCTTCCTAGTCTCCTCGATATTGGGTCTAACCCCCACCTCC
```

```
                2871                                                                                              2952
C12_seq      TGTTAGGCAGAATTCCTTATCTGGTGACACACCCCCATTTCCTGGAGCACATCTCTCCTTCCCAGAACCTCTAAGGTTTGCT
52BH17_.46..
52BH18_.46..
52BH16_.46..
52BH16_.46..
52BH10_.46..
43EH25_.(4..
52BH11_.46..
52BH60_.47.. TGTTAGGCAGAATTCCTTATCTGGTGACACACCCCCATTTCCTGGAGCATCTCTCTCCTTGCCAGAACCTCTAAGGTTTGCT
19H199_04..  TGTTAGGCAGAATTCCTTATCTGGTGACACACCCCCATTTCCTGGAGCATCTCTCTCCTTGCCAGAACCTCTAAGGTTTGCT
```

```
                2953                                                                                              3034
C12_seq      TACGATGGAGCCAGAGGATCCTGGGAGGGAGACTTGGCAGGGGGTGGGAGGGAAGGGGGGATGCGTGACCTGCCCGGT
52BH17_.46..
52BH18_.46..
52BH16_.46..
52BH16_.46..
52BH10_.46..
43EH25_.(4..
52BH11_.46..
52BH60_.47.. TACGATGGAGCCAGAGGATCCTGGGAGGGAGAGACTTGGCAGGGGGTGGGAGGGAAGGGGGGATGGGGTGACCTGCCCGGT
19H199_04..  TACGATGGAGCCAGAGGATCCTGGGAGGGAGAGACTTGGCAGGGGGTGGGAGGGAAGGGGGGATGGGGTGACCTGCCCGGT
```

```
                3035                                                                                              3116
C12_seq      TCTCAGTGGCCACCCTCCGCTACCCTCTCCCAGAACCTGAGCTTGCTGCTCTGACGGCGGCCGTCTGGGTGCGGTTCACTGATCCTGG
52BH17_.46..
52BH18_.46..
52BH16_.46..
52BH16_.46..
52BH10_.46..
43EH25_.(4..
52BH11_.46..
52BH60_.47.. TCTCAGTGGCCACCCTGCGCTACCCTCTCCCAGAACCTGAGCTGCTCTGACGGCGG[CTG]TCTGGGTGCGGTTCACTGATCCTGG
19H199_04..
```

```
                3117                                                                                              3198
C12_seq      TGCTGCAGCTTCCTTACACTTCCAACGAGAGCACAGTTTGGAAAAACAAAAATCAGAATAAGTTGGTCCTGAGTTCTAACT
52BH17_.46..
52BH18_.46..
52BH16_.46..
52BH16_.46..
52BH10_.46..
43EH25_.(4..
52BH11_.46..
52BH60_.47.. TGCTGCAGCTTCCTTACACTTCCAAGAGAGAGAACAGTTTGGAAAAAACAAAAATCAGAATAAGTTGGTCCTGAGTTCTAACT
19H199_04..
```

```
                3199                                                                                              3280
C12_seq      TTGGCTCTTCACCTTTCTAGTCCCCAATTTATATTGTTCCTCCGTGCCTCCAGTTTTACCTCTGAGATAAGGCCAGTAGCCAG
52BH17_.46..
52BH18_.46..
52BH16_.46..
52BH16_.46..
52BH10_.46..
43EH25_.(4..
52BH11_.46..
52BH60_.47.. TTGGCTCTTCACCTTTCTAGTCCCCAATTTATATTGTTCCTCCGTGCCTCCAGTTTTACCTCTGAGATAAGGCCAGTAGCCAG
19H199_04..
```

Section 6.4

```
                3281                                                                                                                              3362
C12_seq         CCCGTCCTTGGCAGGGCTGTTGGTGACAGGAGGGGGTGGGGTGTTCCGTGTGGGAAAAACTCCCTTTGTGACGAATGGTGCGTCCTAGGTGTT
52BH17_46...    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
52BH18_46...    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
52BH16_46...    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
52BH10_46...    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
43EH25 (4...    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
52BH11_46...    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
52BH60_47...    CCCGTCCTTGGCAGGGCTGTGGTGACAGGAGGGGGTGTTCCGTGTGGGAAAACTCCCTTTGTGACGAATGGTGCGTCCTAGGTGTT
19HH99_04...    ----------GGGGGTCCGTGTGGGAAAAACTCCCTTTGTGACGAATGGTGCGTCCTAGGTGTT
................................................................................................

                3363                                                                                                                              3444
C12_seq         CACCAGGTCGTGGCCGCCTCTACTCCCTTTCTTCTTTCTTCCATCCTCTTTTCCTTAAAGAGTCCCCAGTGCTATCTGGGACAT
52BH17_46...    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
52BH18_46...    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
52BH16_46...    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
52BH10_46...    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
43EH25 (4...    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
52BH11_46...    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
52BH60_47...    CACCAGGTCGTGGCCGCCCTCTACTCCCTTTCTTCTTTCTTCCATCCTTCTTTCCTTAAAGAGTCCCCAGTGCTATCTGGGACAT
19HH99_04...    CACCAGGTCGTGGCCCGCCTTCTACTCCCTTTCTCTTTCTTCCATCCTCTTTTCCTTAAAGAGTCCCCAGTGCTATCTGGGACAT
................................................................................................

                3445                                                                                                                              3526
C12_seq         ATTCCTCCGCCCAGACAGAGGTCCCGCTTCCCTAAGGCCCTGCTCTGGGCGTTCTGGGGTTGAGTCCTTGGCAAGCCCAGGAG
52BH17_46...    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
52BH18_46...    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
52BH16_46...    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
52BH10_46...    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
43EH25 (4...    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
52BH60_47...    ATTCCTCCGCCCAGAGACAGGGTCCCGCTTCCCTAAGGCCCTGCTCTGGGCGTTCTGGGGTTGAGTCCTTGGCAAGCCCAGGAG
52BH11_46...    ATTCCTCCGCCCAGACAGAGGTCCCGCTTCCCTAAGGCCCTGCTCTGGGC
19HH99_04...    ATTCCTCCGCCCAGAGACAGGGTCCCGCTTCCCTAAGGCCCTTCTGGGCTTCTGGGGTTGAGTCCTTGGCAAGCCCAGGAG
................................................................................................

                3527                                                                                                                              3608
C12_seq         AGGCGCTCAGGCTTCCCTCGTCCCCCTTCCTCGTCCACCATCTCATCGCCCTGGCTCTCCTGCCCCTTCCCTTACAGGGGTTCC
52BH17_46...    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
52BH18_46...    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
52BH16_46...    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
52BH10_46...    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
43EH25 (4...    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
52BH11_46...    
52BH60_47...    AGGCGCTCAGGCTTCCCTCGTCCCCCTTCCTCGTCCACCATCTCATCGCCCTGGCTCTCCTGCCCCTTCCCTTACAGGGGTTCC
19HH99_04...    AGGCGCTCAGGCTTCCCTCGTCCCCCTTCCTCGTCCACCATCTCATCGCCCTGGCTCTCCTGCCCCTTCCCTACAGGGGTTCC
................................................................................................

                3609                                                                                                                              3690
C12_seq         TGGCTCTGCTCTTCAGACTGAGCCCCGTTCCCCTGCATCCCCGTTCCCCTGCATCCCCCTCCATCCCCCTACAGGAGGCC
52BH17_46...    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
52BH18_46...    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
52BH16_46...    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
52BH10_46...    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
43EH25 (4...    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
52BH11_46...    
52BH60_47...    TGGCTCTGCTCTTCAGACTGAGCCCCGTTCCCCTGCATCCCCGTTCCCCTGCATCCCCCTCCATCCCCCAGAGGCC
19HH99_04...    TGGCTCTGCTCTTCAGACTGAGCCCCGTTCCCCTGCATCCCCGTTCCCCTGCATCCCCCTCCATCCCCCAGAGGCC
................................................................................................
```

```
                3691                                                                                                                              3772
C12_seq         CCAGGCCACCTACTTGGCCTGGACCCCACGAGAGGCCACCCCAGCCCTGTCTACCAGGCTGCCTTTTGGGTGGATTCTCCTC
52BH17_46...    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
52BH18_46...    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
52BH16_46...    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
52BH10_46...    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
43EH25 (4...    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
52BH11_46...    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
52BH60_47...    CCAGGCCACCTACTTGGCCT--------------------
19HH99_04...    CCAGGCCACCTACTTGGCCTGGACCCCACGAGAGGCCACCCCAGCCCTGTCTACCAGGCTGCCTTTTGGGTGGATTCTCCTC
................................................................................................
```

Figure 123: Sanger sequencing alignments of integrated donor DNA in IMR90-4hPOMC-GLuc clone 4. Any mismatched bases are indicated in red (12 pages).

6.4.12 IMR90-4 VAMP2-GLuc clone 11

Sanger sequencing alignment for IMR90-4 VAMP2-GLuc (11) donor DNA insertion. Traces corresponding to dark red sequencing arrows depicted in Figure 65 in the Results section Sanger sequencing of homozygote clones: CQA974: sequencing range #1, CQA975: sequencing range #2, CQA976: sequencing range #3, CQA977: sequencing range #4, CQA965: sequencing range #5, CQA966: sequencing range #6, CQA978: sequencing range #7, CQA979: sequencing range #8 (10 pages).

Page 1 VAMP2-Gluc_sequencing

                          1   82
VAMP2_TEV..... ---TTCAGGTTCCGTCTTCCTTCCACTCCCTTCCTTCCCTTGCTCTTCTGCTGTGTTGCTTGCTCCCCAAGGATGCTCTTTTCCGGAGCA
CQA975_17..... --
CQA974_17..... --CTGCTGTGTTGCTTGCTCCCCAAGGATGCTCTTTTCCGGAGCA
CQA977_17..... --
CQA976_17..... --
CQA966_17..... --
CQA965_17..... --
CQA978_17..... --
CQA979_17..... --
     . . . .

                          83   164
VAMP2_TEV..... CTTCCTTCTCCGGCCTGCACCACGTGATGTCCTCTTGAGCGAGGATCCTCCCCGTGTCTGGGTCCTCCTCCCGGGCATCCTCCTCC
CQA975_17..... CTTCCTTCTCCGGCGCTGCACCACGTGATGTGTCCTCTTGAGCGAGGATCCTCCCCGTGTCTGGGTCCTCCTCCCGGGCATCCTCCTCC
CQA974_17..... --
CQA977_17..... --
CQA976_17..... --
CQA966_17..... --
CQA965_17..... --
CQA978_17..... --
CQA979_17..... --
     . . . .

                          165   246
VAMP2_TEV..... CTCACCCAACCCCATGCCGCTCTTCACTCGCTCTGGGTTCCCTTTTCCTTCTCTCCTTCTCTGGGGCCTGTCCCATCTCTCGTTTCTTA
CQA975_17..... -TCACCCAACCCCATGCCGTCTTCACTCGCTCTGGGTTCCCTTTTCCTTCTCTCCTTCTCTGGGGCCTGTCCCATCTCTCGTTTCTTA
CQA974_17..... CTCACCCAACCCCATGCCGCTCTTCACTCGCTCTGGGTTCCCTTTTCCTTCTCTCCTTCTCTGGGGCCTGTCCCATCTCTCGTTTCTTA
CQA977_17..... --
CQA976_17..... --
CQA966_17..... --
CQA965_17..... --
CQA978_17..... --
CQA979_17..... --
     . . . .

                          247   328
VAMP2_TEV..... GGATGGCCTTCTCCGACGGATGTCTCCCTTGCGTCCGGCCTCCCCTCTTCTTGTAGGCCTGCATCATCACCGTTTTTCTGGACA
CQA975_17..... GGATGGCCTTCTCCGACGGATGTCTCCCTTGCGTCCGGCCTCCCCTCTTCTTGTAGGCCTGCATCATCACCGTTTTTCTGGACA
CQA974_17..... GGATGGCCTTCTCCGACGGATGTCTCCCTTGCGTCCGGCCTCCCCTCTTCTTGTAGGCCTGCATCATCACCGTTTTTCTGGACA
CQA977_17..... --
CQA976_17..... --
CQA966_17..... --
CQA965_17..... --
CQA978_17..... --
CQA979_17..... --
     . . . .

                          329   410
VAMP2_TEV..... ACCCCAAAGTACCCCGTCTCTCCCTGGCTTTAGCCACCTCTCCATCCCTCTTCGTTTCTTCTTGCCTGGACACCCCGTTCTCCTGTG
CQA975_17..... ACCCCAAAGTACCCCGTCTCTCCCTGGCTTTAGCCACCTCTCCATCCCTCTTCGTTTCTTCTTGCCTGGACACCCCGTTCTCCTGTG
CQA974_17..... ACCCCAAAGTACCCCGTCTCTCCCTGGCTTTAGCCACCTCTCCATCCCTCTTCGTTTCTTCTTGCCTGGACACCCCGTTCTCCTGTG
CQA977_17..... --
CQA976_17..... --
CQA966_17..... --
CQA965_17..... --
CQA978_17..... --
CQA979_17..... --
     . . . .

Page 2 VAMP2-Gluc_sequencing

                          411   492
VAMP2_TEV..... GATTCGGGTCACCTCACTCCTTTCATTTGGGCAGCTCCCCCTACCCCCTTACCCTCTCTAGTCTGTGTGCTAGCTCTTCCAGC
CQA975_17..... GATTCGGGTCAACCTCTCACTCCCTTTCATTTGGGCAGCTCCCCCTACCCCCTTACCCTCTCTAGTCTGTGTGCTAGCTCTTCCAGC
CQA974_17..... GATTCGGGTCACCTCACTCCTTTCATTTGGGCAGCTCCCCCTACCCCCTTACCCTCTCTAGTCTGTGTGCTAGCTCTTCCAGC
CQA977_17..... --
CQA976_17..... --
CQA966_17..... --
CQA965_17..... --
CQA978_17..... --
CQA979_17..... --
     . . . .

                          493   574
VAMP2_TEV..... CCCCTGTCATGGGCATCTTCCAGGGGTCCGAGAGCTCAGCTAGTCTTCTTCCTCCAACCCGGGCCCCTATGTCCACTTCAGGA
CQA975_17..... CCCCTGTCATGGGCATCTTCCAGGGGTCCGAGAGCTCAGCTAGTCTTCTTCCTCCAACCCGGGCCCCTATGTCCACTTCAGGA
CQA974_17..... CCCCTGTCATGGGCATCTTCCAGGGGTCCGAGAGCTCAGCTAGTCTTCTTCCTCCAACCCGGGCCCCTATGTCCACTTCAGGA
CQA977_17..... --
CQA976_17..... --
CQA966_17..... --
CQA965_17..... --
CQA978_17..... --
CQA979_17..... --
     . . . .

                          575   656
VAMP2_TEV..... CAGCACATGTTTGCTGCCTCCAGGATCCTGTGTCCCCGAGCTGGGACCACCTTATATATTCCCAGGCGCGGTTAATGTGGCCTG
CQA975_17..... CAGCACATGTTTGCTGCCTCCAGGGATCCTGTGTCCCCGAGCTGGGACCACCTTATATATTCCCAGGCGCGGTTAATGTGGCCTG
CQA974_17..... CAGCACATGTTTGCTGCCTCCAGGATCCTGTGTCCCCGAGCTGGGACCACCTTATATATTCCCAGGCGCGGTTAATGTGGCCTG
CQA977_17..... --
CQA976_17..... --
CQA966_17..... --
CQA965_17..... --
CQA978_17..... --
CQA979_17..... --
     . . . .

                          657   738
VAMP2_TEV..... GTTCTGGGTACTTTATCTGTCCCCTCCACCCCACAGTGGGCGAAGCTTCTAACCTCTTCTCTTCCTCCCACAGGGCCTGA
CQA975_17..... GTTCTGGGTACTTTATCTGTCCCCTCCACCCCACAGTGGGCGAAGCTTCTAACCTCTTCTCTTCCTCCCACAGGGCCTGA
CQA974_17..... GTTCTGGGTACTTTATCTGTCCCCTCCACCCCACAGTGGGCGAAGCTTCTAACCTCTTCTCTTCCTCCCACAGGGCCTGA
CQA977_17..... --
CQA976_17..... --
CQA966_17..... --
CQA965_17..... --
CQA978_17..... --
CQA979_17..... --
     . . . .

                          739   820
VAMP2_TEV..... GAGATCTGGCAGCGGAGAGGGCAGAGAGGAAGTCTTCTAACATGCGTGACGTGGAGGAGAAGAATCCGGCCCTAGGCTCAGATG
CQA975_17..... GAGATCTGGCAGCGGAGAGGGCAGAGAGGAAGTCTTCTAACATGCGTGACGTGGAGGAGAAGAATCCGGCCCTAGGCTCAGATG
CQA974_17..... GAGATCTGGCAGCGGAGAGGGCAGAGAGGAAGTCTTCTAACATGCGTGACGTGGAGGAGAAGAATCCGGCCCTAGGCTCAGATG
CQA977_17..... --
CQA976_17..... --
CQA966_17..... --
CQA965_17..... --
CQA978_17..... --
CQA979_17..... --
     . . . .

241

Section 6.4

```
                     821                                                                                                              902
VAMP2_TEV...  ACCGAGTACAAGACCCACGGTGCGCCTCCGCACACCGGCGAGGGCGGTACGCACCCTCGCCGCGCCGGCGGTTCGCCG
CQA975_17...  ACCGAGTACAAGCCCACCGGTGCGCCTCCGCACCCGGCGAGGACGTCCCAGGGCGCGTACGCACCCTGCCGCCGGTTCGCCG
CQA974_17...  ACCGAGTACAAGCCCACCGGTGCGCCTCCGCACCCGGCGAGGACGTCCCAGGGCGCGTACGCACCCTGCCGCCGGTTCGCCG
CQA977_17...  ACCGAGTACAAGCCCACCGGTGCGCCTCCGCACCCGGCGAGGACGTCCCAGGGCGCGTACGCACCCTGCCGCCGGTTCGCCG
CQA976_17...
CQA966_17...
CQA965_17...
CQA978_17...
CQA979_17...

                     903                                                                                                              984
VAMP2_TEV...  ACTACCCCGCCACGCCACCGCCACACCGTCGATCGGGACGCCACATCGAGCGGGTCACCGAACTCAAGAACTCTTCCTCACGCG
CQA975_17...  ACTACCCCGCCAGGCCGCCACACCGGTCGA------------------------------------------------------
CQA974_17...  CGTCGGGCTCGACATCGGCAAGGTGTTGGTGCGGGACGACGGCCGCCGCCCGGGGTCGGCGGTTCTGGACCACGCCGGAGACCGTCG---
CQA977_17...  ACTACCCCGCCACGCCGCCCACACCGTCGATCGGGACGCCACATCGAGCGGGTCACCGAACTCAAGAACTCTTCCTCACGCG
CQA976_17...
CQA966_17...
CQA965_17...
CQA978_17...
CQA979_17...

                     985                                                                                                              1066
VAMP2_TEV...  CGTCGGGCTCGACATCGGCAAGGTGTTGGTGCGGGACGACGGCCGCCGCCCGGGGTCGGCGGTTCTGGACCACGCCGGAGACCGTCGAA
CQA975_17...
CQA974_17...  CGTCGGGCTCGACATCGGCAAGGTGTTGGTGCGGGACGACGGCCGCCGCCCGGGGTCGGCGGTTCTGGACCACGCCGGAGACCGTCG---
CQA977_17...
CQA976_17...
CQA966_17...
CQA965_17...
CQA978_17...
CQA979_17...

                     1067                                                                                                             1148
VAMP2_TEV...  GCGGGGGGCGGGTTGTTCGCCGAGCATCGGCCGCGGCCGAGTTGAACGGGTTCCCGGCTGGCGCCGGCAGCGAACAGATGGAAG
CQA975_17...
CQA974_17...
CQA977_17...
CQA976_17...
CQA966_17...
CQA965_17...
CQA978_17...
CQA979_17...

                     1149                                                                                                             1230
VAMP2_TEV...  GCCTCCTGGCCCCGCCACCGGCCCAAGAGACCCCGCGTGGTTCCTCGCACCGTCGCCCGACCACCAGGGCAAGGG
CQA975_17...
CQA974_17...
CQA977_17...
CQA976_17...
CQA966_17...
CQA965_17...
CQA978_17...
CQA979_17...
```

```
                     1231                                                                                                            1312
VAMP2_TEV...  TCTGGGCAGCGCGCGTCCGTCCCCGGAGTGGAAGCGGCGGCCGAGCCGCGGCCGGCCCGGGGGTGCCCGCCTTCCTGGAGACCTCCGCGCC
CQA975_17...
CQA974_17...
CQA977_17...
CQA976_17...
CQA966_17...
CQA965_17...
CQA978_17...
CQA979_17...

                     1313                                                                                                            1394
VAMP2_TEV...  CGCAACCTCCCCTTCTTACGAGCGGGCTCGGCGTTCACCGTCACCGCGCAGCGTCGAGGTGCCCGAAGGGACCCGGCCACCTGGTGCA
CQA975_17...
CQA974_17...
CQA977_17...
CQA976_17...
CQA966_17...
CQA965_17...
CQA978_17...
CQA979_17...

                     1395                                                                                                            1476
VAMP2_TEV...  TGACCCGCAAGCCCGGTGCCTGATCTAGAGAGGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCACGC
CQA975_17...
CQA974_17...
CQA977_17...
CQA976_17...
CQA966_17...
CQA965_17...
CQA978_17...
CQA979_17...

                     1477                                                                                                            1558
VAMP2_TEV...  ATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACGTCCTTTCCTTCCTTAATAAAATGAGGAA
CQA975_17...
CQA974_17...  ............TTGACCCTGGAAGGTGCCACTCCCACGTCCTTTCCTTCCTTAATAAAATGAGGAA
CQA977_17...
CQA976_17...
CQA966_17...
CQA965_17...
CQA978_17...
CQA979_17...

                     1559                                                                                                            1640
VAMP2_TEV...  ATTGCATCGCCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGCCAGGACAGCCAAGGGGGAGGATTGGGAAG
CQA975_17...
CQA974_17...  ATTGCATCGCCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGCCAGGACAGCCAAGGGGGAGGATTGGGAAG
CQA977_17...
CQA976_17...
CQA966_17...
CQA965_17...
CQA978_17...
CQA979_17...
```

Section 6.4

```
                    1641                                                                                      1722
VAMP2_TEV...  ACAATAGCAGGCAATGCTGGGGGATGCGGGTGGGGCTCTATGGACTAGTCtagtcaacaccacacggcccctgatcttgtccaactgg
CQA975_17...  -----------------------------------------------------------------------------------------
CQA974_17...  -----------------------------------------------------------------------------------------
CQA977_17...  ACAATAGCCAGGCAATGCTGGGGGATCGCGGTGGGGCTCTATGGACTAGTCACCACCACCGGCCCCCTGATCATCTTGTCCAACTGG
CQA976_17...  CCCTGGATCTTGCTGGGCAAAAGGTCCCACAAGCGGTTGCGCGAGCCACCACTTCTTGAGCAGGTGCAGACACTGACGGTTGGGCAAGCC
CQA966_17...  -----------------------------------AAGGTCGCACAGCGGTTGCGGCACCACCACCTTCTTGAGCAGGTCAGAACAGTTGCAGCC
CQA965_17...  -----------------------------------------------------------------------------------------
CQA978_17...  -----------------------------------------------------------------------------------------
CQA979_17...  -----------------------------------------------------------------------------------------

                    1723                                                                                      1804
VAMP2_TEV...  ccctggatcttgctggcaaaggtcgcacaagcggttgcggcagccaccacttcttgagcaggtcagaacactgacggttggcaagcc
CQA975_17...  -----------------------------------------------------------------------------------------
CQA974_17...  -----------------------------------------------------------------------------------------
CQA977_17...  -----------------------------------------------------------------------------------------
CQA976_17...  -----------------------------------------------------------------------------------------
CQA966_17...  -----------------------------------------------------------------------------------------
CQA965_17...  -----------------------------------------------------------------------------------------
CQA978_17...  -----------------------------------------------------------------------------------------
CQA979_17...  -----------------------------------------------------------------------------------------

                    1805                                                                                      1886
VAMP2_TEV...  ctttgaggcagccagttgtgcagtccacacacagatcgacctgtgcgatgaactgctccatggctccaagtcctgaacc
CQA975_17...  -----------------------------------------------------------------------------------------
CQA974_17...  -----------------------------------------------------------------------------------------
CQA977_17...  CTTTGAGGCAGCCAGTTGTGCAGTCCACACACAGATCGACCTGTGCGATGAACTGCTCCATGGCTCCAAGTCCTTGAACCC
CQA976_17...  CTTTGAGGCACGCCAGTTGTGCCAGTCCACACACAGCATCGACCTGTGCGATGAACTGCTCCAAGTCCTTGAACCC
CQA966_17...  -----------------------------------------------------------------------------------------
CQA965_17...  -----------------------------------------------------------------------------------------
CQA978_17...  -----------------------------------------------------------------------------------------
CQA979_17...  -----------------------------------------------------------------------------------------

                    1887                                                                                      1968
VAMP2_TEV...  aggaatctcaggaatgtcgaacgatcgcctacgcgccgcctgtgcggaactcctttgtccgcctcgttagcgttggcagcgt
CQA975_17...  -----------------------------------------------------------------------------------------
CQA974_17...  -----------------------------------------------------------------------------------------
CQA977_17...  AGGAATCTCAGGAATGTCGACGATCGCCTCGCGCCGCCCTGTGCGGAACTCTTTGTCCGCCTCGTTAGCGTTGGCACCGT
CQA976_17...  AGGAATCTCAGGAATGTCGACGATCGCCTCGCGCCGCCCTGTGCGGAACTCTTTGTCCGCCTTCGTTAGCGTTGGCACCGT
CQA966_17...  -----------------------------------------------------------------------------------------
CQA965_17...  -----------------------------------------------------------------------------------------
CQA978_17...  -----------------------------------------------------------------------------------------
CQA979_17...  -----------------------------------------------------------------------------------------

                    1969                                                                                      2050
VAMP2_TEV...  cctcggatgaactcttcattctgggcgtgcacttgatcgtgggacaggcagatcaagacagcccctggtgcagccagcttcc
CQA975_17...  -----------------------------------------------------------------------------------------
CQA974_17...  -----------------------------------------------------------------------------------------
CQA977_17...  CCTGGGGATGAACTTCTTCATCTTGGGCGTGCACTTGATGTGGGACAGGCAGATCAACAGCCCCTGGTGCAGCCAGCTTTCC
CQA976_17...  CCTGGGGATGAACTTCTTCATCTTGGGCGTGCACTTGATGTGGGACAGGCAGATCAGCAGCCCCTGGTGCAGCCAGCTTTCC
CQA966_17...  CCTGGGGATGAACTTCTTCATCTTGGGCGTGCACTTGATGTGGGACAGGCAGATCAGCAGCCCCTGGTGCAGCCAGCTTTCC
CQA965_17...  CGTTCGTCCACGTCGTCCAGCTCCGACAGCTTCTGGTGCTCGCCTCCAGGACCTGTGTCCACAGTTCACCCTAGTAGTCCACC
CQA978_17...  -----------------------------------------------------------------------------------------
CQA979_17...  -----------------------------------------------------------------------------------------
```

```
                    2051                                                                                      2132
VAMP2_TEV...  gggcaattggcttccatcctcttgagtcaactccagcggcagtcttgccgggcaaacttccgcgcgtcagcatcagaagatcgt
CQA975_17...  -----------------------------------------------------------------------------------------
CQA974_17...  -----------------------------------------------------------------------------------------
CQA977_17...  GGGCAATTGGCTTCCATCTCTCTTTGAGCAACCTCCAGCGGCAGTCTTGCCGGGCAACTTCCCGCGCGTCAGCATCGAGATCCGT
CQA976_17...  GGGCAATTGGCTTCCATCTCTTTGAGCAACCTCCAGCGGCAGTCTTGCCGGGCAACTTCCCGCGCGTCAGCATCGAGATCCGT
CQA966_17...  GGGCATTGGCTTCCATCTCTTTGAGCACTTCCACGCGGCCAGCTTCTTCCCGCGGCGCAACTTCCCGCGCGTCAGCATCGAGATCCGT
CQA965_17...  -----------------------------------------------------------------------------------------
CQA978_17...  -----------------------------------------------------------------------------------------
CQA979_17...  -----------------------------------------------------------------------------------------

                    2133                                                                                      2214
VAMP2_TEV...  ggtcgcgaagttcgtcgcacagccacgagtgttgaaagtcttcgttgttctcggtaggcttcatGAATTCTTGAAAATAT
CQA975_17...  -----------------------------------------------------------------------------------------
CQA974_17...  -----------------------------------------------------------------------------------------
CQA977_17...  GGTCGCGGAAGTTCGTCGGCCACAGCCACGAGTGTTGAAAGTCTTCGTTGTTCTCGGTGGGGGCTTCATGAATTCTTGAAAATAT
CQA976_17...  GGTCGCGAAGTTGCTCGGCCACGGCCACGAGTGTTGAAAGTCTTCGTTGTTCTCGGTGGGGGCTTCATGAATTCTTGAAAATAT
CQA966_17...  -----------------------------------------------------------------------------------------
CQA965_17...  -----------------------------------------------------------------------------------------
CQA978_17...  -----------------------------------------------------------------------------------------
CQA979_17...  -----------------------------------------------------------------------------------------

                    2215                                                                                      2296
VAMP2_TEV...  AAGTTTTCCTTAAGACGCGGTACCAGTGCTGAAGTAAAACTATCATGAATCATGAGATGAGGATGATGGGGCAAAATCACTCCCACAGATGA
CQA975_17...  -----------------------------------------------------------------------------------------
CQA974_17...  -----------------------------------------------------------------------------------------
CQA977_17...  AAGTTTTCCTTAAGACGCGGTACCAGTGCTGAAGTAAAACTATCATGAATCATGAGATGAGGATGATGGGGCAAAATCACTCCCACAGATGA
CQA976_17...  AAGTTTTCCTTAAGACGCGGTACCAGTGCTGAAGTAAAACTATCATGAATCATGAGATGAGGATGATGGGGCAAAATCACTCCCAAGATGA
CQA966_17...  -----------------------------------------------------------------------------------------
CQA965_17...  -----------------------------------------------------------------------------------------
CQA978_17...  -----------------------------------------------------------------------AATCACTCCCAAGATGA
CQA979_17...  -----------------------------------------------------------------------------------------

                    2297                                                                                      2378
VAMP2_TEV...  TCATCATCTTGAGGTTTTTCACCAGTATTTGCGCTTGAGCTTGGCTGCGGCTTGTTTCAAACTGGGAGGCCCCCCGCCTGGAG
CQA975_17...  -----------------------------------------------------------------------------------------
CQA974_17...  -----------------------------------------------------------------------------------------
CQA977_17...  TCATCATCTGAGGTTTTTCACCAGTATTTGCGCTTGAGCTTGGCTGCGGCTTGTTTCAAACTGGGAGGCCCCCCGCCTGGAG
CQA976_17...  TCATCATCTGAGGTTTTTCACCAGTATTTGCGCTTGAGCTTGGCTGCGGCTTGTTTCAAACTGGGAGGCCCCCCGCCTGGAG
CQA966_17...  -----------------------------------------------------------------------------------------
CQA965_17...  TCATCATCTTGAGGTTTTTCACCAGTATTTGCGCTTGAGCTTGGCTGCGGCTTGTTTCAAACTGGGAGGCCCCCCGCCTGGAG
CQA978_17...  -----------------------------------------------------------------------------------------
CQA979_17...  -----------------------------------------------------------------------------------------

                    2379                                                                                      2460
VAMP2_TEV...  TGCATCTGCACGGTGGTCCAGCTCCGACAGCTTCTGGTCGTCCCTCCAGGACCTGTCCACGTTCACCCTCATGATGTCCACC
CQA975_17...  -----------------------------------------------------------------------------------------
CQA974_17...  -----------------------------------------------------------------------------------------
CQA977_17...  TGCATCTGCACGGTGGTCCAGCTCCGACAGCTTCTGGTCGTCCCTCCAGGACCTGTCCACGTTCACCCTCATGATGTCCACC
CQA976_17...  TGCATCTGCACGGTCGTCCAGCTCCGACAGCTTCTGGTCGTCCCTCCAGGACCTGTCCACGTTCACCCTCATGATGTCCACC
CQA966_17...  -----CGGTCGTCCACGTCGTCCAGCTCCGACAGCTTCTGGTCGTCCCTCCAGGACCTGTCCACGTTCACCCTCATGATGTCCACC
CQA965_17...  TGCATCTGCACGGTCGTCCAGCTCCGACAGCTTCTGGTCGTCCCTCCAGGACCTGTCCACGTTCACCCTCATGATGTCCACC
CQA978_17...  -----------------------------------------------------------------------------------------
CQA979_17...  -----------------------------------------------------------------------------------------
```

Section 6.4

```
                           2461                                                                                          2542
VAMP2_TEV...   ACCTCATCCACCTGGCCTGGCTCTGCTCCAGTCTCTGTTACTGGTGAGGTTTGGAGGTTGGAGGGGGTTGCAGGGGGACCACCCTCCC
CQA975_17...   --------------------------------------------------------------------------------------------
CQA974_17...   --------------------------------------------------------------------------------------------
CQA977_17...   ACCTCATCCACCTGGGCCTGGGCTCTGCTCCAGTCTCTGTTACTGGTGAGGTTTGGAGGGGGGTTGCAGGGGGACCACCCTCCC
CQA976_17...   ACCTCATCCACCTGGGCCTGGGCTCTGCTCCAGTCTCTGTTACTGGTGAGGTTTGGAGGGGGGTTGCAGGGGGACCACCCTCCC
CQA966_17...   ACCTCATCCACCTGGGCCTGGGTTCGCTCCGCAGTCTCCTGTTACTGGTGAGGTTTGGAGGGGGGTTGCAGGGGGACCACCCTCCC
CQA965_17...   AAGCTTCACCTGTGTTCTGTCGGCGGCAAAACCCGTTGCGGAAAAAGAAGGTTCACGCGGCGACTACTGCACTTATTATACGGTTCTCCC
CQA978_17...   ACCTCATCCACCTGGGCCTCGCTCCGCAGTCTCCTCCTTTACTGGTGAGGTTGGAGGGGGGTTGCAGGGGGACCACCCTCCC
CQA979_17...   --------------------------------------------------------------------------------------------

................................................................

                           2543                                                                                          2624
VAMP2_TEV...   CAGCCGGGGCAGCAGAGGGGGGGGGGGCCTGGCCAGCCGGGTAGCACATGGTggTTAATTAACGATCTGTAACGGGCCAGAACAGAAAA
CQA975_17...   --------------------------------------------------------------------------------------------
CQA974_17...   --------------------------------------------------------------------------------------------
CQA977_17...   CACCC---
CQA976_17...   CACCCGGGGCAGCAGAGGGGGGGGGCCTGGCCAGCCGGGTAGCACACATGGTGGTTAATTAACGATCTGTAACGGGCCAGAACAGAAAA
CQA966_17...   CACCCGGGGCCAGCAGAGGGGGGGGGGCCTGGCCAGCCGGTAGCACATGGTGGTTAATTAACGATCTGTAACGGGCCAGAACAGAAAA
CQA965_17...   CACCCGGGGCCAGCAGAGGGGGGGGGGGCCTGGCGAGCCGGGTAGCACGACAGAGACATGGTGGTTAATTAACGATCTGTAACGGGCCAGAACAGAAAA
CQA978_17...   --------------------------------------------------------------------------------------------
CQA979_17...   --------------------------------------------------------------------------------------------

................................................................

                           2625                                                                                          2706
VAMP2_TEV...   CGAAACAAAGACGTAGAGTTGAGCAAGCAGAGGGTCAGGCAAAAGCCGTGGAGAGACGCCGGCTGAGTCTAGGTAGGCTCCAAGGGAGC
CQA975_17...   --------------------------------------------------------------------------------------------
CQA974_17...   --------------------------------------------------------------------------------------------
CQA977_17...   --------------------------------------------------------------------------------------------
CQA976_17...   CGAAACAAAGACGTAGAGTTGAGCAAGCAGAGGGTCAGGCAAAAGCCGTGGAGAGACGCCGGCTGAGTCTAGGTAGGCTCCAAGGGAGC
CQA966_17...   CGAAACAAAGACGTAGAGTTGAGCAAGCAGAGGGTCAGGCAAAAGCCGTGGAGAGACGCCGGCTGAGTCTAGGTAGGCTCCAAGGGAGC
CQA965_17...   CGAAACAAAGACGTAGAGTTGAGCAAGCAGAGGGTCAGGCAAAAGCCGTGGAGAGACGCCGGCTGAGTCTAGGTAGGCTCCAAGGGAGC
CQA978_17...   --------------------------------------------------------------------------------------------
CQA979_17...   --------------------------------------------------------------------------------------------

................................................................

                           2707                                                                                          2788
VAMP2_TEV...   GCCGGACAAAGGCCCGGTCTCGACCTGAGCTTTAAACTTACCTAGACGGCGGAGCCGACCAGTTCAGGAGCCACCACAGGCGGGAG
CQA975_17...   --------------------------------------------------------------------------------------------
CQA974_17...   --------------------------------------------------------------------------------------------
CQA977_17...   --------------------------------------------------------------------------------------------
CQA976_17...   GCCGGACAAAGGCCCGGTCTCGACCTGAGCTTTAAACTTACCTAGACGGCGGAGCCGACCAGTTCAGGAGCCACCACAGGCGGGAG
CQA966_17...   GCCGGACAAAGGCCCGGGTCTCGACCTGAGCTTTAAACTTACCTAGACGGCGGAGCCGACCAGTTCAGGAGCCACCACAGGCGGGAG
CQA965_17...   GCCGGACAAAGGCCCGGGTCTCGACCTGAGCTTTAAACTTACCTAGACGGCGGAGCCGACCAGTTCAGGAGCCACCACAGGCGGGAG
CQA978_17...   GACAGAAAAGCCCATCCTTAGGCCTCTCCTCTTCCTAGTCTCCTGATAATGGGTCTAACCCCCACCCTCCTGTTAGGCAGATT
CQA979_17...   ----ACCTCCTGTTAGGCAGATT

................................................................

                           2789                                                                                          2870
VAMP2_TEV...   GCGGCAGAACCGACTCAACCGGCGTGGATGGCGCGCGCGGCGCGGCGCGTGAAGGAGAGGAGATGCGAGCCCCTCG
CQA975_17...   --------------------------------------------------------------------------------------------
CQA974_17...   --------------------------------------------------------------------------------------------
CQA977_17...   --------------------------------------------------------------------------------------------
CQA976_17...   --------------------------------------------------------------------------------------------
CQA966_17...   GCGGCAGAGAAGCGCACTCAACCGCGCGTGGATGGGCGGCGGCGGCCTCAGGTAGGGCGCGCGGCGCGTGAAGGAGAGAGAGATGCGAGCCCCTCG
CQA965_17...   GCGGCAGAGAACCGGACTCAACCGGCGTGGATGGGCGGGCGGCGCGTGAAGGAGAGAGAGAGAGATGCGAGCCCCTCG
CQA978_17...   --------------------------------------------------------------------------------------------
CQA979_17...   --------------------------------------------------------------------------------------------

................................................................
```

```
                           2871                                                                                          2952
VAMP2_TEV...   AAGCTTCAGCTGTGTTCTGTGGCGGCAAAACCCGTTGCGGAAAAAGAAGGTTCACGCGGCGACTACTGCACTTATTATACGGTTCTCCC
CQA975_17...   --------------------------------------------------------------------------------------------
CQA974_17...   --------------------------------------------------------------------------------------------
CQA977_17...   --------------------------------------------------------------------------------------------
CQA976_17...   AAGCTTCAGCTGTGTTCTGTGGCGGCGGCAAAACCCGTTGCGGAAAAAGAAGGTTCACGCGGCGACTACTGCACTTATTATACGGTTCTCCC
CQA966_17...   AAGCTTCAGCTGTGTTCTGTGGCGGCAAACCCGTTGCGGAAAAAGAAGGTTCACGCGGCGACTACTGCACTTATTATACGGTTCTCCC
CQA965_17...   --------------------------------------------------------------------------------------------
CQA978_17...   --------------------------------------------------------------------------------------------TC
CQA979_17...   --------------------------------------------------------------------------------------------

................................................................

                           2953                                                                                          3034
VAMP2_TEV...   CCACCCTCGGGAAAAAGGCGGGAGGCCAGTACAGACACATCACTTTCCCACGTTTACCCGCGGCCACCTTCTCTAGGCACCGGTTC
CQA975_17...   --------------------------------------------------------------------------------------------
CQA974_17...   --------------------------------------------------------------------------------------------
CQA977_17...   --------------------------------------------------------------------------------------------
CQA976_17...   CCACCCTCGGGAAAAAGGCGGGAGGCCAGTACAGACACATCACTTTCCCACGTTTACCCGCGGCCACCTTCTCTAGGCACCGGTTC
CQA966_17...   CCACCCTCGGGAAAAAGGCGGGAGGCCAGTACAGACACATCACTTTCCCACGTTTACCCGCGGCCACCTTCTCTAGGCACCGGTTC
CQA965_17...   --------------------------------------------------------------------------------------------
CQA978_17...   --------------------------------------------------------------------------------------------
CQA979_17...   --------------------------------------------------------------------------------------------

................................................................

                           3035                                                                                          3116
VAMP2_TEV...   AAITGCCGACCCCTCCCCCCCAACTTCTCGGGGACTGTGGGCGATGTGGGCTCTGCCCACTGAGGTCGGACAGTACTAAGCTTT
CQA975_17...   --------------------------------------------------------------------------------------------
CQA974_17...   --------------------------------------------------------------------------------------------
CQA977_17...   --------------------------------------------------------------------------------------------
CQA976_17...   --------------------------------------------------------------------------------------------
CQA966_17...   AAITGCCGACCCCTCCCCCCCAACTTCTCTCGG-------------------------------------------------------------
CQA965_17...   AAITGCCGACCCCTCCCCCCCAACTTCTCTCGG---------------------------------------------------------------
CQA978_17...   AAITGCCGACCCCTCCCCCCCAACTTCTCGGGGACTGTGGGCGATGTGGGCTCTGCCCACTGAGGTCGGACAGTACTAAGCTTT
CQA979_17...   --------------------------------------------------------------------------------------------

................................................................

                           3117                                                                                          3198
VAMP2_TEV...   GACAGAAAAGCCCCATCCTTAGGCCTCTCCTCTTCCTAGTCTCCTGATAATGGGTCTAACCCCCACCCTCCTGTTAGGCAGATT
CQA975_17...   --------------------------------------------------------------------------------------------
CQA974_17...   --------------------------------------------------------------------------------------------
CQA977_17...   --------------------------------------------------------------------------------------------
CQA976_17...   GACAGAAAAGCCCCATCCTTAGGCCTCTCCTCTTCCTAGTCTCCTGATAATGGGTCTAACCCCCACCCTCCTGTTAGGCAGATT
CQA966_17...   GACAGAAAAGCCCCATCCTTAGGCCTCTCCTCTTCCTAGTCTCCTGATAATGGGTCTAACCCCCACCCTCCTGTTAGGCAGATT
CQA965_17...   --------------------------------------------------------------------------------------------
CQA978_17...   GACAGAAAAGCCCCATCCTTAGGCCTCTCCTCTTCCTAGTCTCCTGATAATGGGTCTAACCCCCACCCTCCTGTTAGGCAGATT
CQA979_17...   --------------------------------------------------------ACCTCCTGTTAGGCAGATT

................................................................

                           3199                                                                                          3280
VAMP2_TEV...   CCTTATCTGGTGACACACCCCATTTCCTGGACCATCTCTCTCTCCTTGCCAGAACCTCTAAGGTTTGCTTACGATGGAGCCA
CQA975_17...   --------------------------------------------------------------------------------------------
CQA974_17...   --------------------------------------------------------------------------------------------
CQA977_17...   --------------------------------------------------------------------------------------------
CQA976_17...   CCTTATCTGGTGACACACCCCATTTCCTGGACCATCTCTCTCTCCTTGCCAGAACCTCTAAGGTTTGCTTACGATGGAGCCA
CQA966_17...   CCTTATCTGGTGACACACCCCATTTCCTGGAGGCCATCTCTCTCTCTTGCCAGA------------------------------
CQA965_17...   --------------------------------------------------------------------------------------------
CQA978_17...   CCTTATCTGGTGACACACCCCATTTCCTGGAGGCCATCTCTCTCTCCTTGCCAGAACCTCTAAGGTTTGCTTACGATGGAGCCA
CQA979_17...   CCTTATCTGGTGACACACCCCATTTCCTGGAGGCCATCTCTCTCTCCTTGCCAGAACCTCTAAGGTTTGCTTACGATGGAGCCA

................................................................
```

Section 6.4

Page 9 alignment blocks (VAMP2_TEV and CQA974–CQA979 / CQA965, CQA966, CQA976, CQA977, CQA978, CQA979 _17 samples):

3281
VAMP2_TEV... GAGAGGATCCTGGGGAGGAGGAGAGAGCTTGGCAGGGGGGTGGGAGAGGGGAAGGGGGGGAGGGGATCGCGTGACCTGCCCGGTTCTCAGTGGCCAC 3362
CQA975_17...
CQA974_17...
CQA977_17...
CQA976_17...
CQA966_17...
CQA965_17...   ----GAGAGGATCCTGGGGAGGAGGAGAGCTTG----
CQA978_17...   GAGAGGATCCTGGGGAGGAGGAGAGCTTG----
CQA979_17...   GAGAGGATCCTGGGGAGGAGGAGAGCTTGGCAGGGGGGTGGGAGAGGGGAAGGGGGGGAGGGGATCGCGTGACCTGCCCGGTTCTCAGTGGCCAC

3363
VAMP2_TEV... CCTGCGCTACCCTCTTCCCAGAACCTGACCTGCTCTGACGCGCCGTCTCGGTGCCGTTTCACTGATCCTGGTGCTGCAGCTTCC 3444
CQA975_17...
...
CQA979_17... CCTGCGCTACCCTCTTCCCAGAACCTGACCTGCTCTGACGCGGG[●]GTCTGGTGCGTTTCACTGATCCTGGTGCTGCAGCTTCC

3445
VAMP2_TEV... TTACACTTCCCAAGAGAGAAGCAGTTTGGAAAAACAAAATCAGAATAAAGTTGGTCCTCAGTTCTAACTTTGGCTCTTCACC 3526
CQA975_17...
...
CQA979_17... TTACACTTCCCAAGAGAGAAGCAGTTTGGAAAAACAAAATCAGAATAAAGTTGGTCCTCAGTTCTAACTTTGGCTCTTCACC

3527
VAMP2_TEV... TTTCTAGTCCCCAATTTATATATGTTCCTCCCGTCGGTCAGTTTTTACCTGTGAGATAAGGCCAGTAGCCACCCCGTCCTGGCA 3608
...
CQA979_17... TTTCTAGTCCCCAATTTATATATGTTCCTCCCGTCGGTCAGTTTTTACCTGTGAGATAAGGCCAGTAGCCACCCCGTCCTGGCA

3609
VAMP2_TEV... GGGCTGTGTGAGGAGGGGGTGTTCCGTTGTGAGAAAACTCCCTTTTGTTGAGAATGGTGCGTCCTAGGTGTTCACCAGGTCGTGG 3690
...
CQA979_17... GGGCTGTGTGAGGAGGGGGTGTTCCGTTGTGAGAAAACTCCCTTTTGTTGAGAATGGTGCGTCCTAGGTGTTCACCAGGTCGTGG

Page 10 alignment blocks:

3691
VAMP2_TEV... CCGCCTCTACTCCCTTTCTCTTCTTTCCTCCATCCTTCTTTCTTCCTTAAAGAGTCCCCAGTGCTATCTGGACATATTCCTCCGCCCA 3772
...
CQA979_17... CCGCCTCTACTCCCTTTCTCTTCTTTCCTCCATCCTTCTTTCTTCCTTAAAGAGTCCCCAGTGCTATCTGGACATATTCCTCCGCCCA

3773
VAMP2_TEV... GAGCAGAGGTCCCGCTTCCCTAAGGCCCTGCTCTGGGCTTCTGGGGTTTGAGTCCTTGGCAAGCCCAGGAGAGGGGCTCAGGCT 3854
...
CQA979_17... GAGCAGAGGTCCCGCTTCCCTAAGGCCCTGCTCTGGGCTTCTGGGGTTTGAGTCCTTGGCAAGCCCAGGAGAGGGGCTCAGGCT

3855
VAMP2_TEV... TCCCTGTCCCCTTCCTCTGTCCACCATCTCATGCCCCTGGCTCTCCTGCCCCTTCCCTACAGGGGGTTCCTGGCTCTGCTCTT 3936
...
CQA979_17... TCCCTGTCCCCTTCCTCTGTCCACCATCTCATGCCCCTGGCTCTCCTGCCCCTTCCCTACAGGGGGTTCCTGGCTCTGCTCTT

3937
VAMP2_TEV... CAGACTGAGCCCCGGTTCCCCTGCATCCCGTTCCCCTGCATCCCCAGAGCCCCAGGCCCACCTAC 4018
...
CQA979_17... CAGACTGAGCCCCGGTTCCCCTGCATCCCGTTCCCCTGCATCCCCAGAGCCCCAGGCCCACCTAC

4019
VAMP2_TEV... TTGGCCTGGACCCCACGACGAGGCCACCCCCAGCCCTGTCTACCAGGCTGCCCTTTTGGGTGGATTCTCCTC 4087
...
CQA979_17... TTGGCCTGGACCCCACGACGAGGCCACCCCC----

Figure 124: Sanger sequencing alignments of integrated donor DNA in IMR90-4 VAMP2-GLuc clone 11. Any mismatched bases are indicated in red (10 pages).

245

Section 6.4

6.5 Insert confirmation optimization

In order to find the best annealing temperature with which to run the insert confirmation PCR, a temperature gradient was run with a range of annealing temperatures from 63 to 53 °C and several primer combinations. The PCR products amplified with the following primer pairs were run on an agarose gel: A) Fpcr_5'ofHA + AAVS1xHA_R2, B) Fpcr_5'ofHA + AAVS1xHA_R3, C) 5'ProbeWT_F + AAVS1xHA_R2, D) 5'ProbeWT_F + AAVS1xHA_R3 (Figure 125). Those samples labeled with upper-case letters were amplified from non-transfected SIMA gDNA, with an expected product size of approximately 1600 bp. Those samples labeled with lower-case letters were amplified from SIMA hPOMC-GLuc clone 5 with an expected insertion at the AAVS1 safe harbor locus, increasing the expected product size to approximately 3700 bp. Unexpectedly, there was no amplified product from neither gDNA source nor any primer combination.
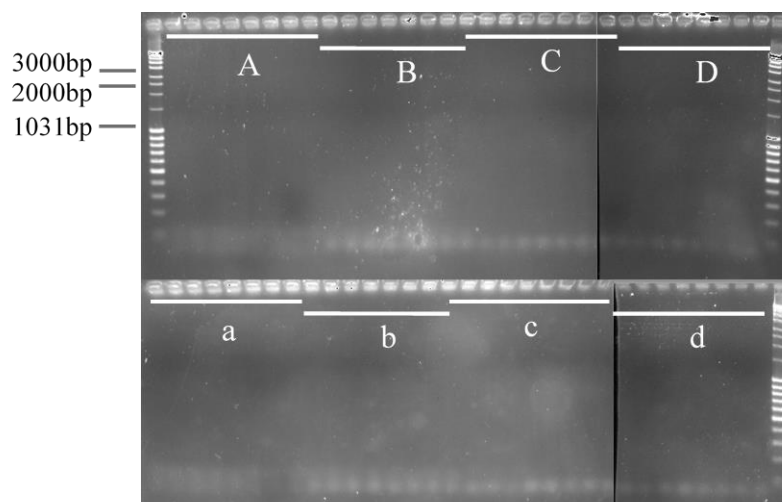
Figure 125: Temperature gradient of primer pairs designed to amplify integrated donor DNA region, including the homology arm sequence. Primer pairs: A) Fpcr_5'ofHA + AAVS1xHA_R2, B) Fpcr_5'ofHA + AAVS1xHA_R3, C) 5'ProbeWT_F + AAVS1xHA_R2, D) 5'ProbeWT_F + AAVS1xHA_R3. Areas marked in upper-case letters amplified with non-transfected SIMA gDNA, areas marked in lower-case letters amplified with SIMA hPOMC-GLuc clone 5 gDNA with expected insert.

In order to troubleshoot the problems with the WT AAVS1 safe harbor locus PCR amplification, the positive control homology arm primers (Fpcr-803 and Rpcr-wt-183) from a previously published optimization of CRISPR/Cas9 cloning[74] were tested. The expected product size of WT gDNA when amplifying with these primers was 1400 bp. The products of this PCR, run using the standard thermocycling program can be seen in the agarose gel A (Figure 126). The same master mix was also run using the thermocycling program

246

recommended by Oceguera *et al*[74] and the products were separated on an agarose gel seen in segment B (Figure 126). Neither thermocycling program successfully amplified the area around the AAVS1 safe harbor locus.
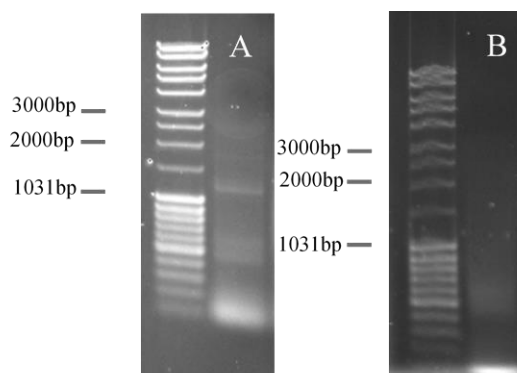
Figure 126: PCR for insert confirmation using primers FPCR-803 + Rpcr-wt-183, primers used in Oceguera et al[74] to confirm insertion of donor DNA at AAVS1 safe harbor locus. WT product expected at 1400bp. Gel A) amplification of non-transfected SIMA gDNA using standard cycling program. Gel B) amplification of non-transfected SIMA gDNA using cycling program from Oceguera et al, with the following changes from standard: 10 second denaturation and 1 minute elongation steps.

Troubleshooting was continued with the Oceguera primers, however no product could be amplified with DMSO or without DMSO, corresponding respectively to the A and B labels (Figure 127). There was also no difference in amplification whether the PCR was run with an annealing temperature at 53 °C or 62 °C.
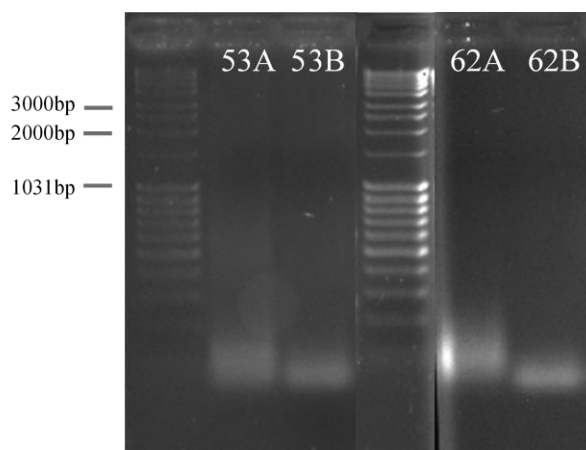
Figure 127: Modified PCR for insert confirmation using primers FPCR-803 + Rpcr-wt-183, primers used in Oceguera et al to confirm insertion of donor DNA at AAVS1 safe harbor locus. WT product expected at 1400bp. The annealing step of the PCR was run at either 53°C or 62°C with standard master mix (A with DMSO or B without DMSO). No product was detected with any protocol.

Troubleshooting of the AAVS1 safe harbor locus amplification continued by testing a wide ranging annealing temperature gradient from 72 – 48 °C with the Oceguera primers including DMSO (A) and excluding DMSO (B) in gDNA from non-transfected SIMA cells (Figure 128). A new reverse primer was included in the middle gel panel: Rpcr-wt-3'HA was found in the middle of the right homology arm sequence (panels C and D). In combination with

247

the previously tested Fpcr-5'ofHA primer, PCR products of various sizes were amplified both with (C) and without (D) DMSO. The correct WT amplification size for this primer pair was 850 bp. The two highest annealing temperatures in combination with the master mix including DMSO produced specific products at 850 bp. This protocol was next used to amplify both gDNA from non transfected SIMA cells (E) and gDNA from SIMA hPOMC-GLuc clone 5 (F) in a temperature gradient with annealing temperatures ranging from 72 – 55 °C. The highest annealing temperatures once again produced specific products at 850 bp with the non-transfected SIMA gDNA (E). At the highest annealing temperatures with the SIMA hPOMC-GLuc gDNA, an additional band was amplified at 3000 bp, exactly the size of the product including the hPOMC-GLuc donor DNA. Interestingly, the WT band at 850 bp was also still amplified in the clone DNA. This indicates that there was probably one WT allele in this clone as well as one allele in which the donor DNA has successfully integrated, making it a homozygous clone (Figure 128).
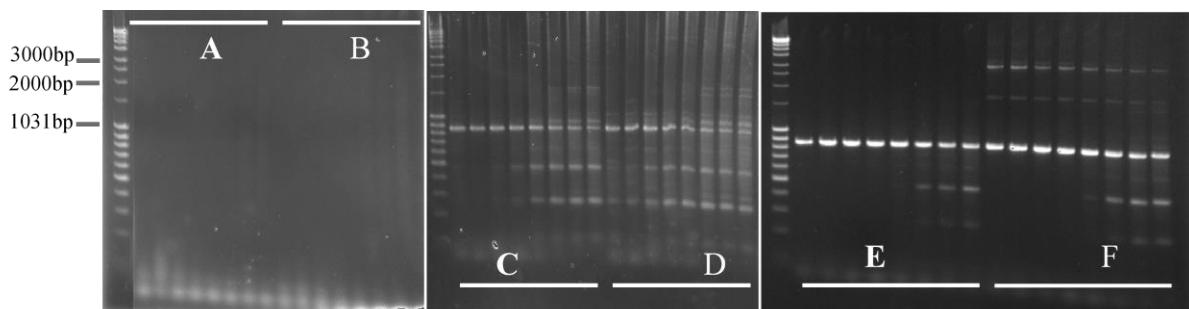
Figure 128: Series of temperature gradient PCRs to troubleshoot insert confirmation amplification. Temperature gradient PCR (72-48°C) PCR for primers FPCR-803 + Rpcr-wt-183, primers used in Oceguera et al. WT product expected at 1400bp. The standard master mix prepared with (A) and without (B) DMSO. Temperature gradient using new primers Fpcr-5'ofHA + Rpcr-wt-3'HA with the expected WT product size of 850bp, the standard master mix prepared with (C) and without DMSO (D). Temperature gradient (72-55°C) using primers 5'Probe_F + Rpcr-wt-3'HA and standard master mix with DMSO. E) non-transfected SIMA gDNA, F) SIMA hPOMC-GLuc clone 5 with putative insert (expected band at 3000bp).

The insert confirmation protocol was further optimized by using the Phusion polymerase in place of the DreamTAQ polymerase. These optimization steps resulted in a more precise and specific amplification of the insert (wells 1-5 and 7) and the WT/insert pair as seen in well 6 (Figure 129).
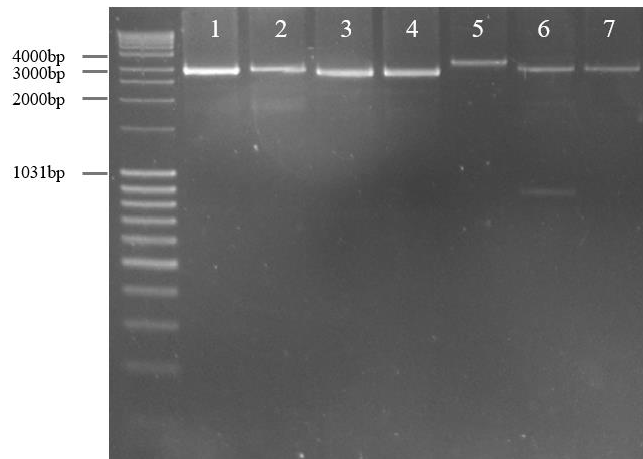
Figure 129: Optimization of the insert confirmation PCR with 5'Probe_F + Rpcr-wt-3'HA primers with various clones used in this project. Optimization of the protocol was completed with the standard Phusion master mix and cycling on the Bio-Rad thermal cyclers, resulting in a more precise and specific insert and insert/WT band combination.

6.6 Optimization - Analysis of off-target integration events

- Southern blot
- Ligation-mediated PCR

6.6.1 Southern blot

Test blots

The first Southern blot test was run with the SIMA Random-Insertion_hPOMC1-26GLucprototype clone gDNA and gDNA from SIMA hPOMC-GLuc clone 2. Each gDNA sample was digested with HindIII and VspI, with expected digested DNA fragment sizes in the SIMA hPOMC-GLuc sample of 1.8 kbp and 36 kbp, respectively. The membrane was hybridized with the GLucF1R1 probe at 60 °C. Despite running the gel overnight for 15 hours, the separation of the larger DNA fragments was not great enough to make VspI a useful restriction enzyme to use in this test. The largest fragment in the DNA ladder was at 21 kbp and it migrated the same distance as all the largest DNA fragments (Figure 130). Absolutely no band was seen in either digestion of the prototype (lanes 2A and 2B). However, a faint band was seen at the expected size of hPOMC-GLuc digested with HindIII (in lane 2A) as indicated by the white arrow. Another faint band can be seen at approximately 10 kbp when SIMA hPOMC-GLuc was digested with VspI (lane 2B). This would be an unexpected product, however, since the VspI digestion was expected to lay above the 21 kbp marker, as indicated by the second white arrow (Figure 130).
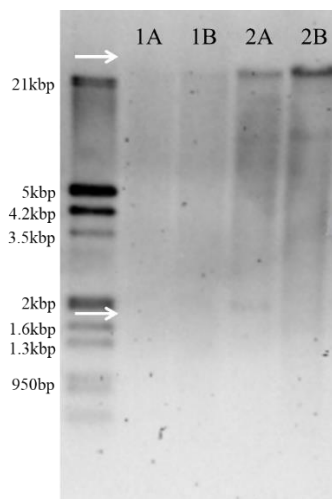
Figure 130: Southern blot with GLuc-F1R1 probe (510 bp) at 60 °C. 1 = SIMA Random-Insertion_hPOMC1-26GLuc prototype gDNA, 2 = SIMA hPOMC-GLuc gDNA, A = HindIII digestion, B = VspI digestion. Expected product sizes indicated with white arrows at 36 kbp and 1.8 kbp.

Due to the large size of the VspI digested fragment, this restriction enzyme was no longer used in future Southern blot digestions. Instead, the enzymes BglI and EcoRI were added to the test, which were expected to produce fragment sizes of 1.5 kbp and 10.3 kbp, respectively. A new gel was run, once again with the SIMA Random-Insertion_hPOMC1-26GLuc prototype gDNA, as well as with non-transfected SIMA gDNA and two different SIMA hPOMC-GLuc gDNA digests. gDNA digests were labelled as non-transfected DNA (WT), SIMA Random-Insertion_hPOMC1-26GLuc prototype DNA (1), SIMA hPOMC-GLuc clone 14 (2), SIMA hPOMC-GLuc clone 29 (3), and plasmid (Pl) (Figure 131). Restriction enzymes used to digest the gDNA were labelled as: BglI (A), EcoRI (B), HindIII (C). White arrows indicate the expected sizes of the digested hPOMC-GLuc fragments. The membrane was hybridized with the GLucF2R2 probe (which is 395 bp long) at 48 °C. As expected, no band was detected in the WT lane. Once again, absolutely no band was detected in the SIMA prototype digests. Very faint bands were also seen in the SIMA hPOMC-GLuc clone 14 digestion (2), this time all three different restriction enzyme digestions were at the expected size. The second SIMA hPOMC-GLuc clone digestions (clone 29, lanes labelled 3) had extremely strong bands at the expected sizes as well as multiple unexpected bands at multiple sizes for each digest. The cut plasmid was included in this gel, which was stained with a strong band at 6.6 kbp, as expected (Figure 131).

Section 6.6
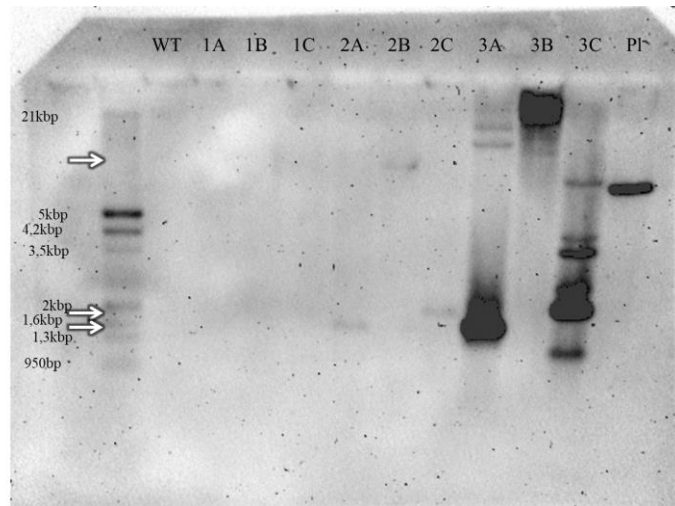
Figure 131: Southern blot with GLuc-F2R2 probe at 48 °C. WT = non-transfected SIMA gDNA, 1 = SIMA Random-Insertion_hPOMC1-26GLuc prototype gDNA, 2 = SIMA hPOMC-GLuc clone 14 gDNA, 3 = SIMA hPOMC-GLuc clone 29 gDNA, A = BglI digestion (1.5 kbp), B = EcoRI digestion (10.3 kbp), C = HindIII digestion (1.8 kbp). Expected product sizes indicated with white arrows.

Modifications to probe size, annealing temperature, and stringency wash temperature

Minor changes in probe visualization could be seen when the hybridization temperature is changed. The membrane was hybridized with the GLucF2R2 probe at 52 °C rather than 48 °C and the intensity of the bands is quite similar, but the background signal of the membrane is much higher (Figure 132).
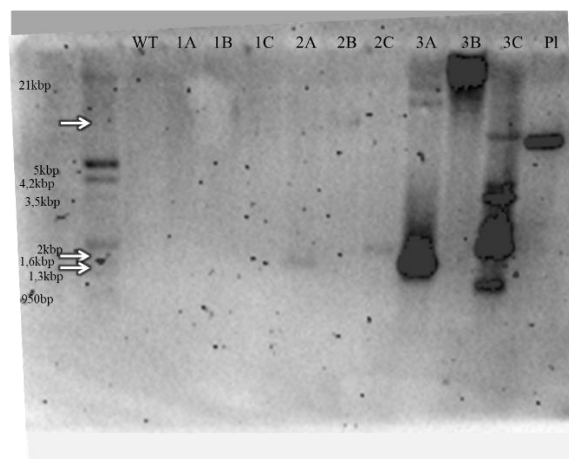
Figure 132: Southern blot with GLuc-F2R2 probe at 52 °C. WT = non-transfected SIMA gDNA, 1 = SIMA Random-Insertion_hPOMC1-26GLuc prototype gDNA, 2 = SIMA hPOMC-GLuc clone 14 gDNA, 3 = SIMA hPOMC-GLuc clone 29 gDNA, A = BglI digestion (1.5 kbp), B = EcoRI digestion (10.3 kbp), C = HindIII digestion (1.8 kbp). Expected product sizes indicated with white arrows.

A similar pattern of bands was seen on the membrane hybridized with the probe GLucF2R4, which is 221 bp long, at 53 °C (Figure 133). However the bands were even fainter than those hybridized with GLucF2R2 (Figure 131 and Figure 132). The use of this smaller probe improved neither the quality nor the utility of this detection method.

251

Section 6.6
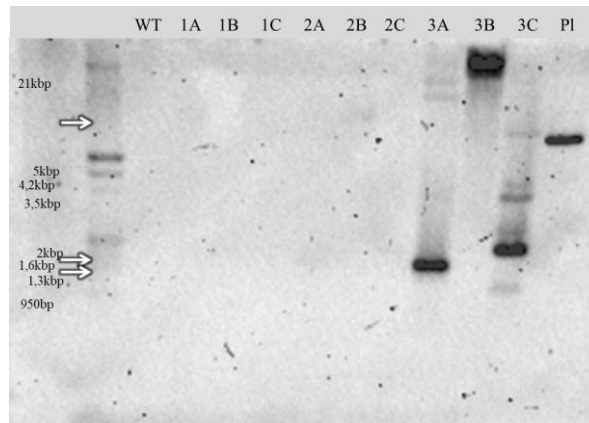
Figure 133: Southern blot with GLuc-F2R4 probe at 53 °C. WT = non-transfected SIMA gDNA, 1 = SIMA Random-Insertion_hPOMC1-26GLuc prototype gDNA, 2 = SIMA hPOMC-GLuc clone 14 gDNA, 3 = SIMA hPOMC-GLuc clone 29 gDNA, A = BglI digestion (1.5 kbp), B = EcoRI digestion (10.3 kbp), C = HindIII digestion (1.8 kbp). Expected product sizes indicated with white arrows.

The same membrane was hybridized with the GLucF3R3 probe, which is 145 bp long, at 49 °C (Figure 134). Once again, the intensity of the bands detected with the even smaller probe GLucF3R3 was too faint to make this a reliable method.
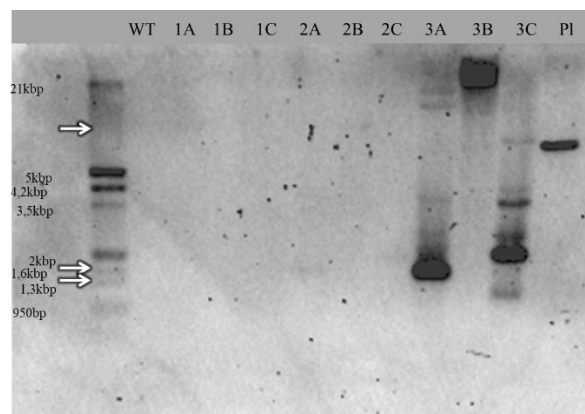
Figure 134: Southern blot with GLuc-F3R3 probe at 49 °C. WT = non-transfected SIMA gDNA, 1 = SIMA Random-Insertion_hPOMC1-26GLuc prototype gDNA, 2 = SIMA hPOMC-GLuc clone 14 gDNA, 3 = SIMA hPOMC-GLuc clone 29 gDNA, A = BglI digestion (1.5 kbp), B = EcoRI digestion (10.3 kbp), C = HindIII digestion (1.8 kbp). Expected product sizes indicated with white arrows.

The probe GLucF3R4, which is 299 bp long, was used for hybridization at 50 °C (Figure 135). The size of this probe was also smaller than the original GLucF2R2 probe and the bands of the first hPOMC-GLuc clone were too faint to reliably identify.

Section 6.6
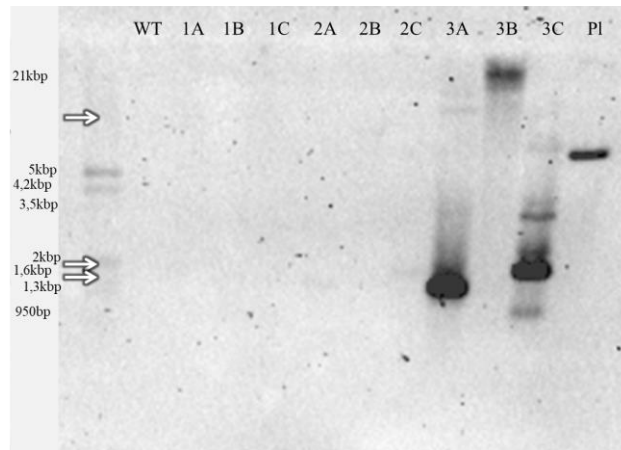
Figure 135: Southern blot with GLuc-F3R4 probe at 50 °C. WT = non-transfected SIMA Random-Insertion_hPOMC1-26GLuc prototype gDNA, 2 = SIMA hPOMC-GLuc clone 14 gDNA, 3 = SIMA hPOMC-GLuc clone 29 gDNA, A = BglI digestion (1.5 kbp), B = EcoRI digestion (10.3 kbp), C = HindIII digestion (1.8 kbp). Expected product sizes indicated with white arrows.

A hybridization at 58 °C with the puromycin probe, which detects the puromycin resistance sequence in the donor DNA, had a similar pattern as the smaller GLuc probes (Figure 136). There were absolutely no bands seen for the SIMA Random-Insertion_hPOMC1-26GLuc prototype DNA, the bands in the hPOMC-GLuc clone 14 digests were barely visible, and the bands in the hPOMC-GLuc clone 29 digests were strong and occur at many different sizes. With this probe it was more evident that the EcoRI digest in the second hPOMC-GLuc clone (lane 3B) did not have the expected 10.3 kbp band, which was surprising, since the bands at the expected sizes of BglI (1.5 kbp) and HindIII (1.8 kbp) were so intense.
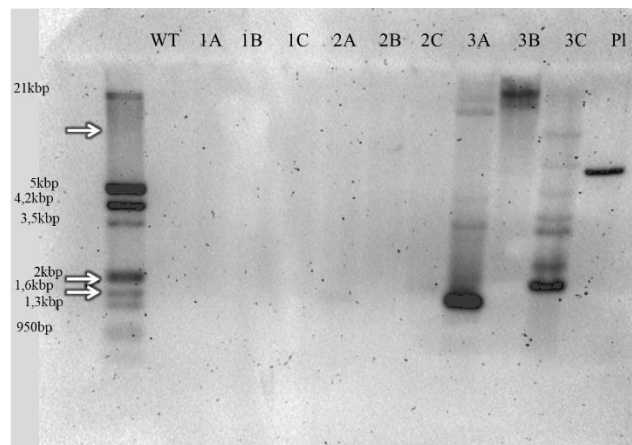
Figure 136: Southern blot with puromycin probe at 58 °C. WT = non-transfected SIMA gDNA, 1 = SIMA Random-Insertion_hPOMC1-26GLuc prototype gDNA, 2 = SIMA hPOMC-GLuc clone 14 gDNA, 3 = SIMA hPOMC-GLuc clone 29 gDNA, A = BglI digestion (1.5 kbp), B = EcoRI digestion (10.3 kbp), C = HindIII digestion (1.8 kbp). Expected product sizes indicated with white arrows.

A hybridization was carried out with a newly digested gDNA of three different SIMA hPOMC-GLuc clones transferred to a new membrane, in order to determine if the problem with the faint bands was due to the test method or due to the clone. The digested gDNA layout in the new blot was as follows: non-transfected SIMA gDNA (WT), SIMA hPOMC-GLuc clone 18

(1), SIMA hPOMC-GLuc clone 19 (2), SIMA hPOMC-GLuc clone 23 (3), and plasmid (Pl). Restriction enzymes used to digest the gDNA were labelled as: BglI (A), EcoRI (B), HindIII (C). The samples 3C and 2C were pipetted out of order, so the labels were highlighted in red. Among other experimental conditions, a hybridization at 50 °C of a probe that binds to the Ef1-HTLV promoter was tested with this membrane. In the digests of all the hPOMC-GLuc clones, very faint bands were seen at the expected sizes of the EcoRI and HindIII digestions. However, in the first clone no band was seen in the BglI digestion, and the second and third clones appeared to have the faint band at around 6.6 kbp as opposed to the expected 1.5 kbp (Figure 137).
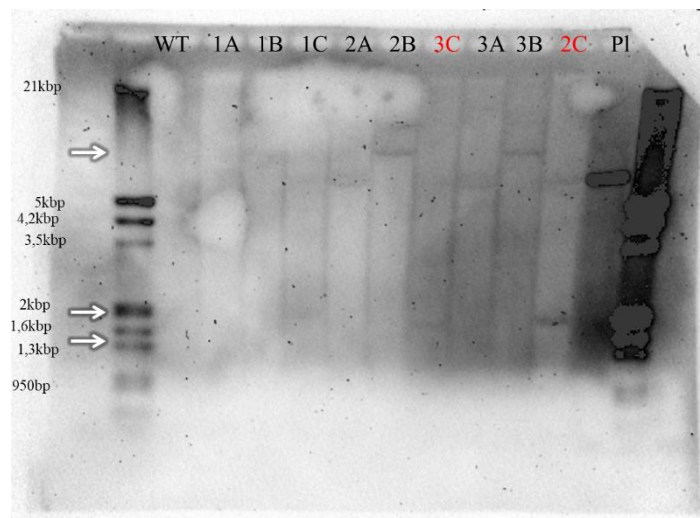
Figure 137: Southern blot with Ef1a-HTLV probe at 50 °C. WT = non-transfected SIMA DNA, 1 = SIMA hPOMC-GLuc clone 18, 2 = SIMA hPOMC-GLuc clone 19, 3 = SIMA hPOMC-GLuc clone 23, Pl = plasmid (Pl), A = BglI digestion (1.5 kbp), B = EcoRI digestion (10.3 kbp), C = HindIII digestion (1.8 kbp). Expected product sizes indicated with white arrows.

The membrane of Figure 138 was partially cut off due to the extreme overexposure of the right half of the membrane. This blot was prepared in order to test the Roche membrane. The digested gDNA layout was as follows: SIMA hPOMC-GLuc clone 18 (1), SIMA hPOMC-GLuc clone 19 (2), BglI digestion (A), EcoRI digestion (B), HindIII digestion (C). The use of the probe detecting only the HTLV segment of the promoter, hybridized at 50 °C, did not result in any bands of the expected sizes. Furthermore the use of the more sensitive CSD star detection reagent appeared to have increased the background of the blot, rather than the sensitivity to detect the probed DNA fragments (Figure 138).

Section 6.6
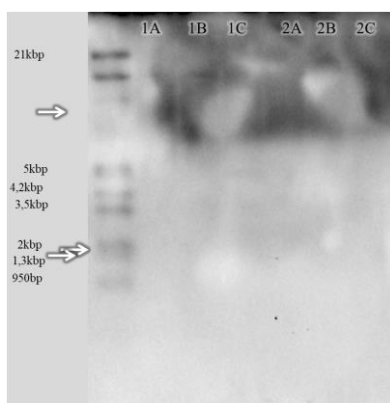
Figure 138: Southern blot with HTLV probe at 50 °C, Roche membrane, CSDstar detection solution. 1 = SIMA hPOMC-GLuc clone 18, 2 = SIMA hPOMC-GLuc clone 19, A = BglI digestion (1.5 kbp), B = EcoRI digestion (10.3 kbp), C = HindIII digestion (1.8 kbp). Expected product sizes indicated with white arrows.

Despite a great deal of testing, optimization, and troubleshooting, it was unfortunately impossible to develop a reliable and robust working protocol to use a DIG-based Southern blot system to detect on-target and possible off-target integrations of donor DNA in genetically modified clones.

6.6.2 Ligation-mediated PCR

Expected PCR products can be deduced from the genomic sequence surrounding the AAVS1 safe harbor locus. If the digestion and ligation of the adapters had been successful, then the smallest expected PCR product from the HindIII digestion/ligation would be at 3500 bp, from the BspHI digestion/ligation at 7000 bp, and from the AseI digestion/ligation at 9000 bp. A standard temperature gradient PCR protocol was run in order to find any possible annealing temperatures which might correctly amplify any product from the ligation reaction. None of these potential products were successfully amplified (Figure 139). The PCR products from the HindIII digestion/ligation, which were separated on an agarose gel, are found in wells marked A (amplification with DMSO) and B (amplification without DMSO). The PCR products from the BspHI digestion/ligation, which were separated on an agarose gel, are found in wells marked C (with DMSO) and D (without DMSO). The PCR products from the AseI digestion/ligation, where were separated on an agarose gel, are found in wells marked E (with DMSO) and F (without DSMO).
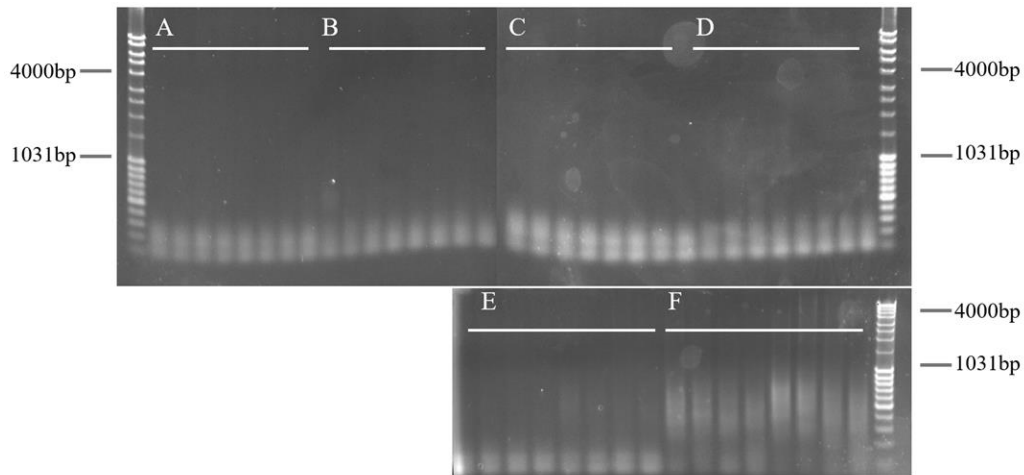
Figure 139: Gel electrophoresis of ligation-mediated adapter PCR. Temperature gradient with annealing temperatures ranging from 67 – 50 °C. PCR of the HindIII digestion/ligation, amplification with DMSO (A) and amplification without DMSO (B). PCR of the BspHI digestion/ligation, amplification with DMSO (C) and without DMSO (D). PCR of the AseI digestion/ligation, amplified with DMSO (E) and without DSMO (F).

6.6.3   Optimization - Cellular localization of Gaussia Luciferase

6.6.3.1   Differential fractionation

The use of differential fractionation to identify the localization of a protein takes advantage of the ability to use multiple centrifugation steps, each at increasing velocities, to separate different components of homogenized cells. The three main components separated in this experiment were the cellular debris, the cytosolic fraction, and the membrane fraction. Contrary to expectations, the highest amount of luciferase was by far found in both the cellular debris and the cytosolic fraction, with hardly any luciferase detected in the membrane fraction (Figure 140).

**Luciferase Activity
Differential Fractionation**
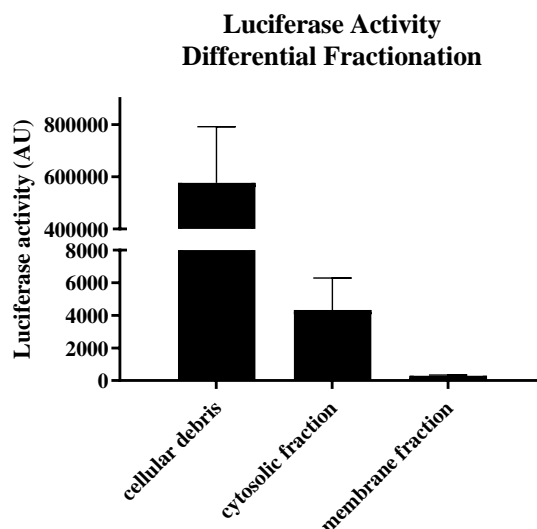
Figure 140: Luciferase activity in subcellular fractions of IMR90-4 hPOMC-GLuc 6, as determined by cellular homogenization and differential centrifugation.

6.6.3.2 Immunofluorescence Antibody Validation

The following antibodies were tested, but not used further in this project due to underperformance.

- Rabbit-α-Gaussia luciferase (Santa Cruz)
- Mouse-α-Golgi97
- Golgi97-Leptin HPA
- Rabbit-α-GAPDH
- Mouse-α-Islet1 (1B1 / ab86501)

Gaussia luciferase

The NEB rabbit-α-GLuc antibody was tested in an IMR90-4 hPOMC-GLuc clone as well as with non-transfected IMR90-4 cells, to prove the specificity of both the GLuc antibody and the secondary antibody. Two dilutions were prepared for both cell types: 1:500 (Figure 141) and 1:1000 (Figure 48). The 1:500 dilution resulted in slightly higher background staining in WT cells, therefore subsequent experiments were performed using the 1:1000 dilution of the GLuc antibody.
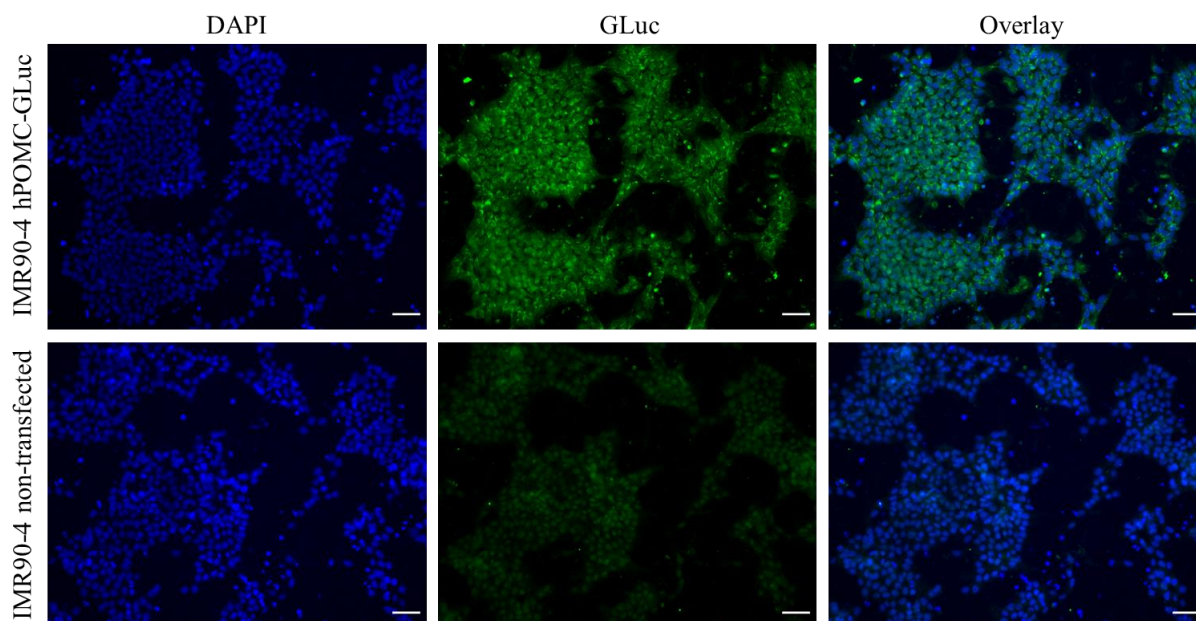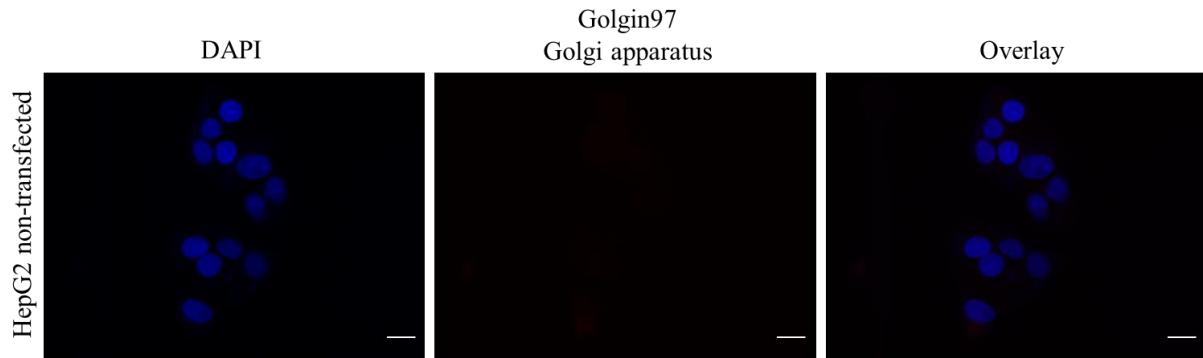
Figure 141: IMR90-4 hPOMC-GLuc clone 6 stained with IF indirect labelling method to detect GLuc. Cells were fixed, permeabilized and then incubated overnight with a 1:500 dilution of rabbit-α-GLuc (NEB) primary antibody. Secondary antibody incubation with Alexa-Fluor α-rabbit 488 diluted to 1:750. Cells were additionally incubated with DAPI for nuclei staining. scale bar = 50µm

The rabbit-α-GLuc (Santa Cruz) primary antibody was tested at a 1:50 dilution (Figure 142) in IMR90-4 hPOMC-GLuc clone 6. In section B only very faint staining was seen, indicating the incompatibility of this antibody with IF. This antibody was not tested or used any further.
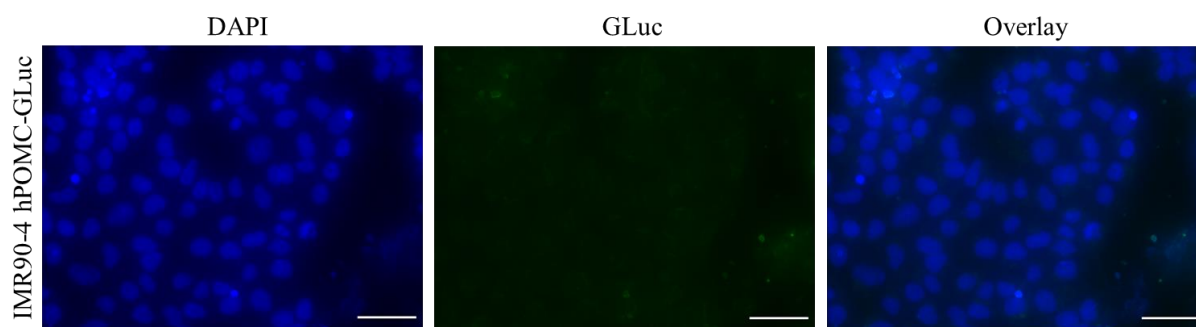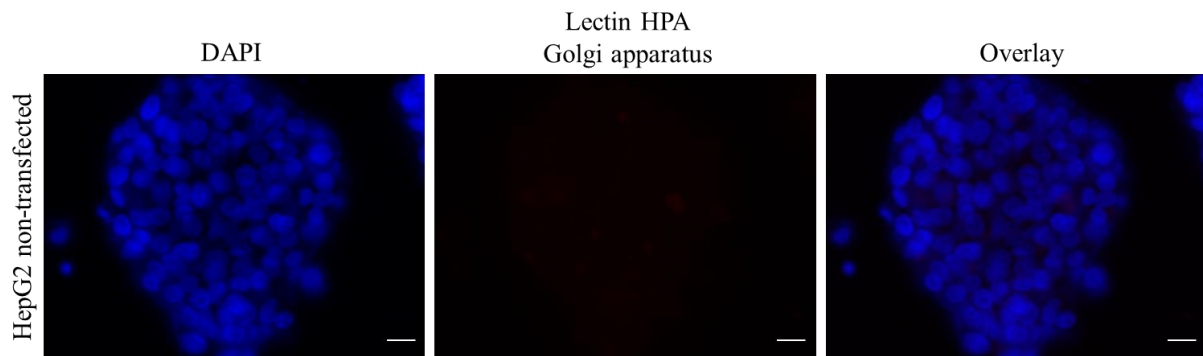

Figure 142: IMR90-4 hPOMC-GLuc clone 6 stained with IF indirect labelling method to detect GLuc. Cells were fixed, permeabilized and then incubated overnight with a 1:50 dilution of rabbit-α-GLuc (Santa Cruz) primary antibody. Secondary antibody incubation with Alexa-Fluor α-rabbit 488 diluted to 1:750. Cells were additionally incubated with DAPI for nuclei staining. scale bar = 25µm

Golgi apparatus

Golgin97 is a protein constituting a part of the *trans*-Golgi network. The Golgin97 antibody was tested with multiple dilutions (1:50, 1:500, 1:750, and 1:1000) and with 2 different secondary antibodies (anti-mouse Alexa-532 and anti-mouse Alexa 680). None of the dilutions with either of the secondary antibodies resulted in successful staining of the *trans*-Golgi network. An example image series demonstrating the failed staining is seen in Figure 143, with

258

the 1:50 dilution of the Golgin97 antibody. The antibody concentrations with higher dilution factors were equally unsuccessful in visualizing the Golgi apparatus.
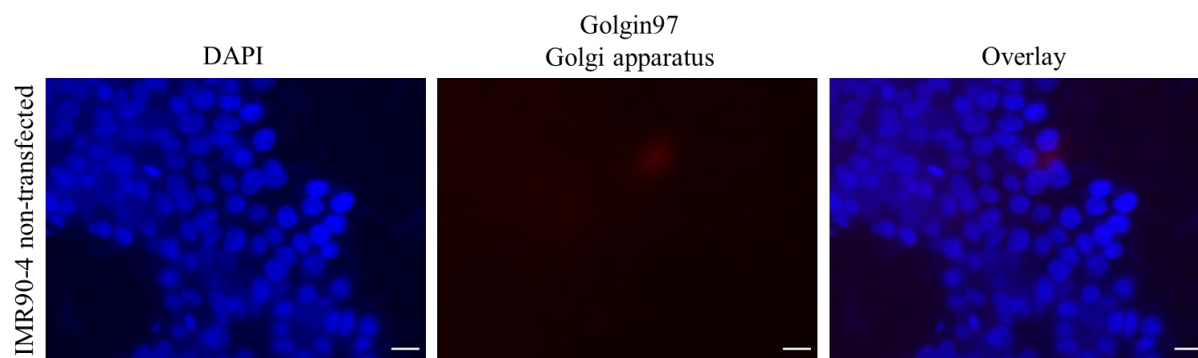
Figure 143: Non-transfected IMR90-4 stained with IF indirect labelling method to detect Golgin-97. Cells were fixed, permeabilized and then incubated overnight with a 1:50 dilution of mouse-α-Golgin-97 primary antibody. Secondary antibody incubation with Alexa-Fluor α-mouse 532 diluted to 1:1000. Cells were additionally incubated with DAPI for nuclei staining. scale bar = 10μm

HepG2 cells were used to troubleshoot the staining of Golgi97 by the modifying the permeabilization steps of the immunofluorescence protocol. The testing of these treatments was run in parallel with a replacement Golgi identification marker, Lectin Helix pomatia agglutinin (HPA) which selectively binds to α-N-acetylgalactosamine residues, an intermediate sugar found on serine and threonine residues transferring between the *cis*-Golgi and the *trans*-Golgi (Thermo Fisher Cat# L32454). The Lectin HPA is conjugated to Alexa 647 and does not need to undergo secondary antibody incubation. The optimization scheme is summarized in Table 56.

Table 56: Optimization scenario of Lectin HPA and trans-Golgi network staining of Golgin97 in HepG2 cells. PFA = paraformaldehyde, ON = overnight, min = minute

Treatment ID:	A1	A2	A3	A4	A5	A6
fixation:	4% PFA	4% PFA	4% PFA	4% PFA	4% PFA	4% PFA
permeabilization:	0.05% tritonX for 10 min	0.25% tritonX for 20 min	0.1% tritonX for 5 min	0.01% tritonX for 5 min	0.05% tritonX for 10 min	0.25% tritonX for 20 min
blocking:	60min 10%FCS, 1%BSA, 0.01% TritonX	60min 10%FCS, 1%BSA, 0.01% TritonX	60min 10%FCS, 1%BSA, 0.01% TritonX	60min 10%FCS, 1%BSA, 0.01% TritonX	60min 10%FCS, 1%BSA, 0.01% TritonX	60min 10%FCS, 1%BSA, 0.01% TritonX
incubation:	ON 4°C Lectin HPA (1:20)	ON 4°C Lectin HPA (1:20)	ON 4°C golgin97	ON 4°C golgin97	ON 4°C golgin97	ON 4°C golgin97

The optimization of the permeabilization step did not result in the identification of Golgi97 through IF in HepG2 cells. An example image series shows the failed staining under extended permeabilization standards (treatment A6, Figure 144). The remaining changes to the permeabilization conditions also did not result in improved staining (data not shown).

Figure 144: HepG2 cells stained with IF indirect labelling method to detect Golgin-97. Cells were fixed, permeabilized and then incubated overnight with a 1:50 dilution of mouse-α-Golgin-97 primary antibody. Secondary antibody incubation with Alexa-Fluor α-mouse 532 diluted to 1:1000. Cells were additionally incubated with DAPI for nuclei staining. scale bar = 10μm

Furthermore, none of the conditions tested from Table 56 with a replacement Lectin HPA marker were able to visualize the Golgi apparatus. Figure 145 exemplifies, with the A2 treatment, the failed Lectin HPA staining.

Figure 145: HepG2 cells stained with IF indirect labelling method to detect Golgi apparatus. Cells were fixed, permeabilized and then incubated overnight with a 1:20 dilution of Lectin HPA Golgi stain. Cells were additionally incubated with DAPI for nuclei staining. scale bar = 10μm

An optimized protocol for the labelling of Golgin-97 was communicated by Dr, Michael Krauss. This IF protocol finally resulted in stained Golgi apparatus in HepG2 cells, revealing the typical accumulation of the organelle in the perinuclear centrosomal region[110] (Figure 146, top panel). However, parallel staining of the same antibody preparation and protocol resulted in much weaker staining in non-transfected SIMA cells (middle panel), and non-transfected IMR90-4 cells (bottom panel).

260
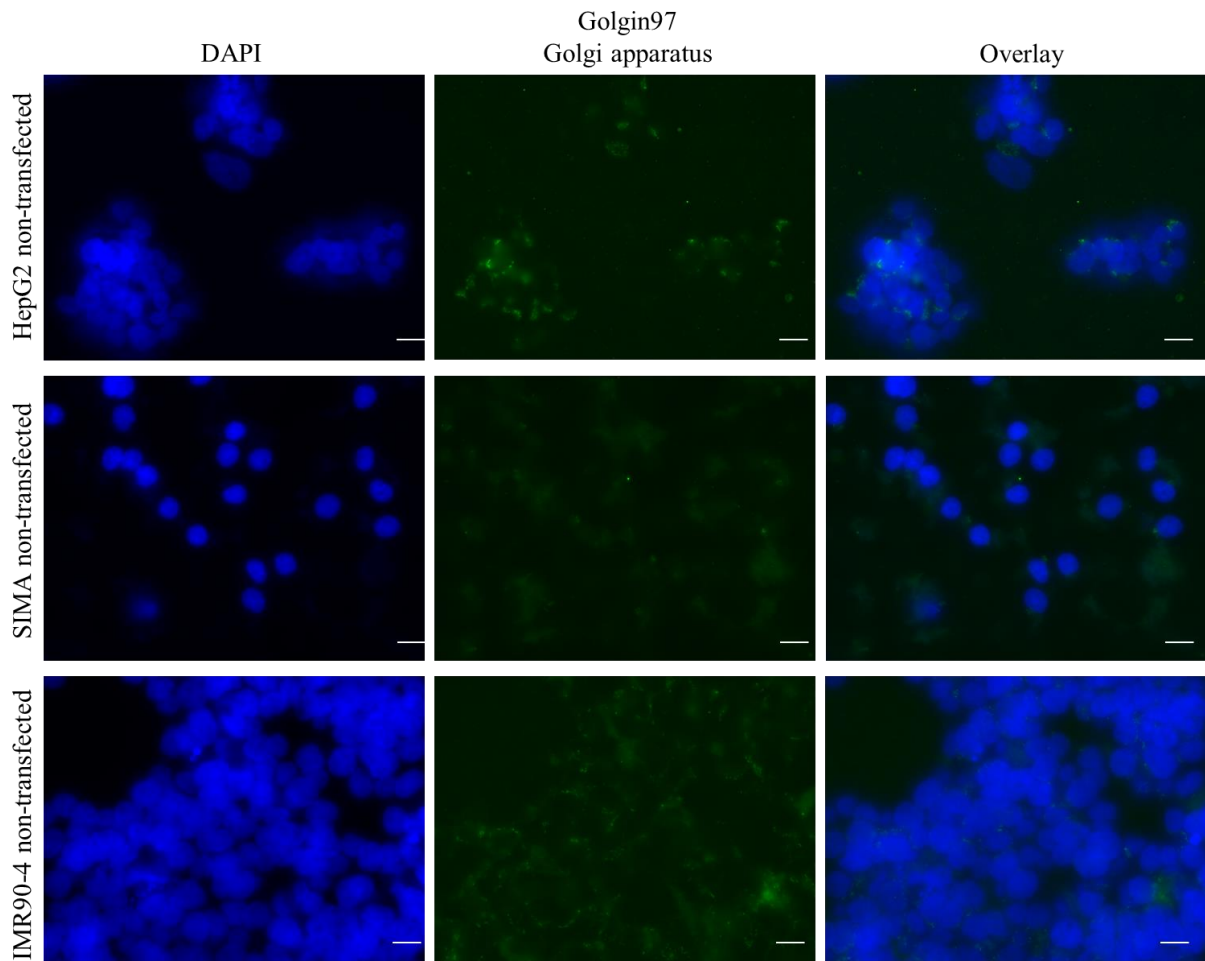
Figure 146: HepG2, SIMA, and IMR90-4 cells stained with IF indirect labelling method to detect Golgin-97 using alternative protocol provided by Dr. Michael Krauss. Cells were fixed, permeabilized and then incubated overnight with a 1:50 dilution of mouse-α-Golgin-97 primary antibody. Secondary antibody incubation with Alexa-Fluor α-mouse 532 diluted to 1:1000. Cells were additionally incubated with DAPI for nuclei staining. scale bar = 10μm

Cytosol

In order to establish a cytosolic marker, the antibody against GAPDH was tested in two concentrations, both a 1:50 dilution (Figure 147) and a 1:250 dilution (not shown). Neither condition stained successfully and this marker was not considered in any subsequent analyses.
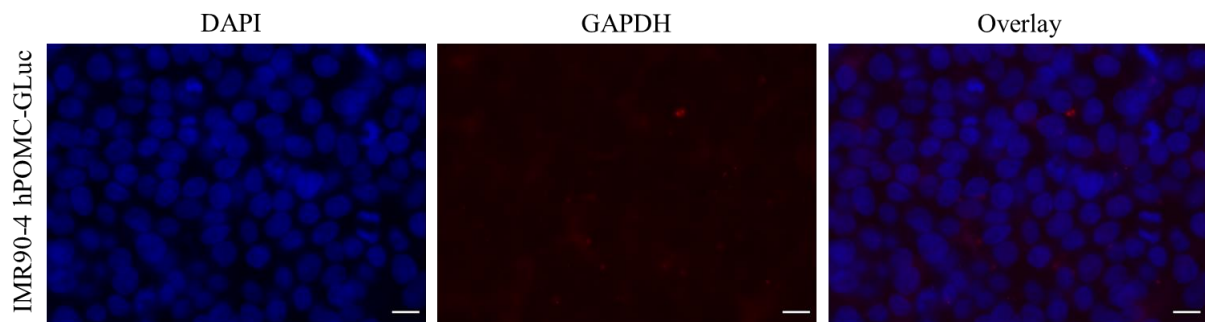
Figure 147: IMR90-4 hPOMC-GLuc clone 6 stained with IF indirect labelling method to detect GAPDH. Cells were fixed, permeabilized and then incubated overnight with a 1:50 dilution of mouse-α-GAPDH primary antibody. Secondary antibody incubation with Alexa-Fluor α-mouse 532 diluted to 1:1000. Cells were additionally incubated with DAPI for nuclei staining. scale bar = 10μm

Motor neurons

The mouse-α.Islet1 [1B1] antibody was tested in the IMR90-4 hPOMC-GLuc clone 6, differentiated according to Maury *et al* for 30 days, at a dilution of 1:500 (within the manufacturer's suggested range). The visualization of the labeled protein resulted in a larger percentage of stained cells than expected (Figure 148). Therefore, the same protocol was carried out in undifferentiated IMR90-4 VAMP-GLuc clone 11, which should not be labeled with Islet1. However, these undifferentiated cells were also distinctly labeled with the antibody (). This indicates unspecific staining of the antibody, the manufacturer was contacted and a replacement antibody was received.
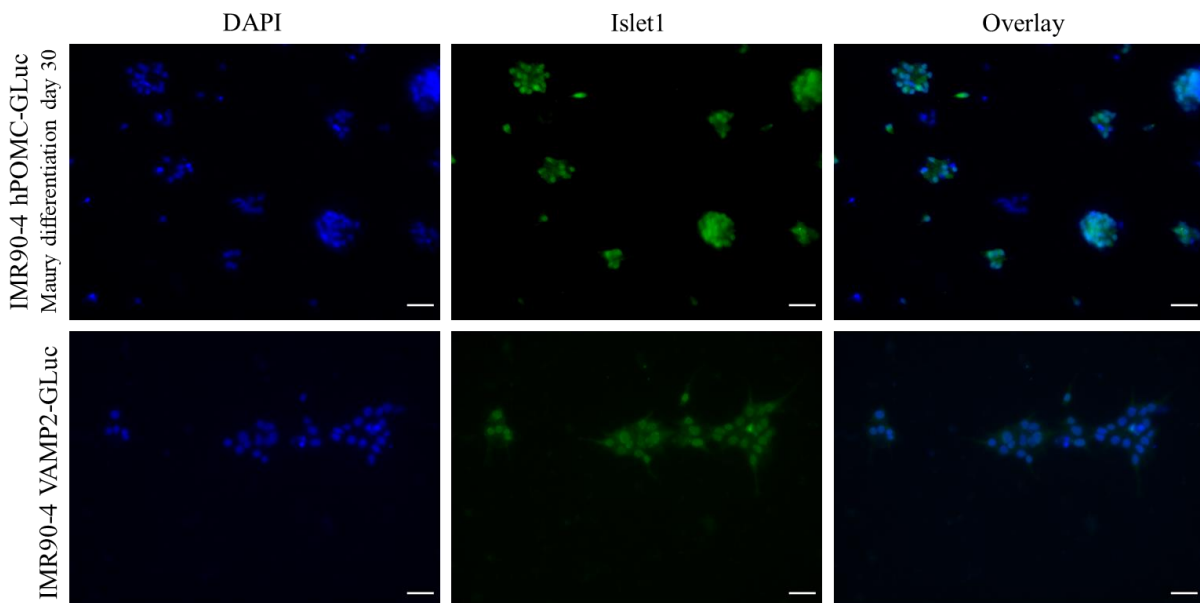
Figure 148: IMR90-4 hPOMC-GLuc clone 6 (Maury D30) and IMR90-4 VAMP2-GLuc stained with IF indirect labelling method to detect Islet1. Cells were fixed, permeabilized and then incubated overnight with a 1:500 dilution of mouse-α-Islet1 [1B1] primary antibody. Secondary antibody incubation with Alexa-Fluor α-mouse 532 diluted to 1:1000. Cells were additionally incubated with DAPI for nuclei staining. scale bar = 50µm

6.7   Expression analysis in non-transfected IMR90-4 cells

The relative gene expression of GLuc, *HB9*, *ISLET1*, and *CHAT* were compared between non-transfected IMR90-4 cells and IMR90-4 hPOMC-GLuc cells. The expression of GLuc in hPOMC-GLuc cells was significantly higher than that in the non-transfected cells, in which no GLuc could be detected. There was no statistical difference in the trace expression of *HB9*, *ISLET1*, and *CHAT* between non-transfected IMR90-4 cells and IMR90-4 hPOMC-GLuc cells (Figure 149).

Change in expression of GLuc and key genes important for motor neuron differentiation in non-transfected IMR90-4 and IMR90-4 hPOMC-GLuc
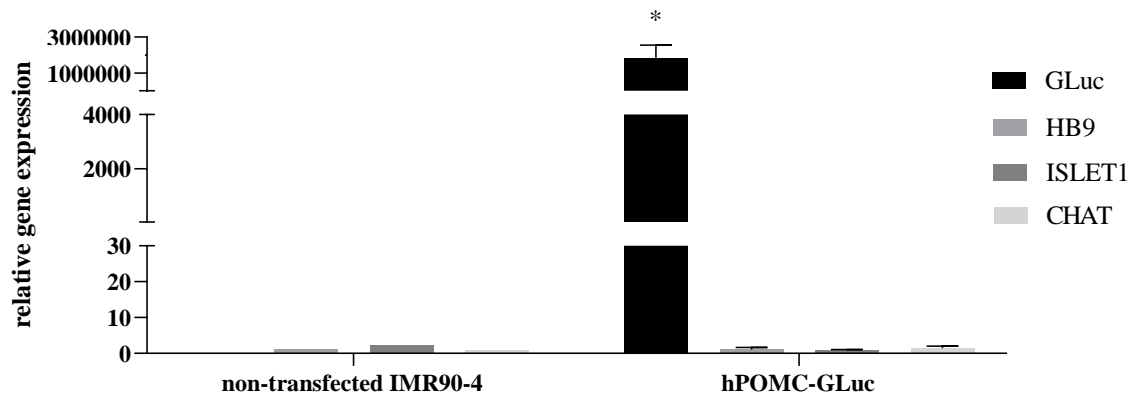
Figure 149: GLuc expression in undifferentiated non-transfected IMR90-4 cells and hPOMC-GLuc clone 6. Gene expression of GLuc and selected genes associated with MN status were calculated by the ΔΔCt method, using the genes RPS23 and PPIA used for normalization. Statistical significance measured by t-test, change in expression levels between differentiated clone versus clone in pluripotent state. Statistical significance evaluated with t-test * = p < 0.05. Non-transfected IMR90-4 measurements consisted of one biological replicate. hPOMC-GLuc clone measurements consisted of 2-6 biological replicates.

Acknowledgements

Statement of authorship

I hereby certify that this thesis has been composed by me, Brit-Maren Schjeide, and is based on my own independent work, unless stated otherwise. No other person's work has been used without due acknowledgement in this thesis. All references and verbatim extracts have been quoted, and all sources of information, including graphs and data sets, have been specifically acknowledged. I further declare that I have not submitted this thesis at any other institution in order to obtain a degree.

Brit-Maren Schjeide   Berlin, 21.04.2021