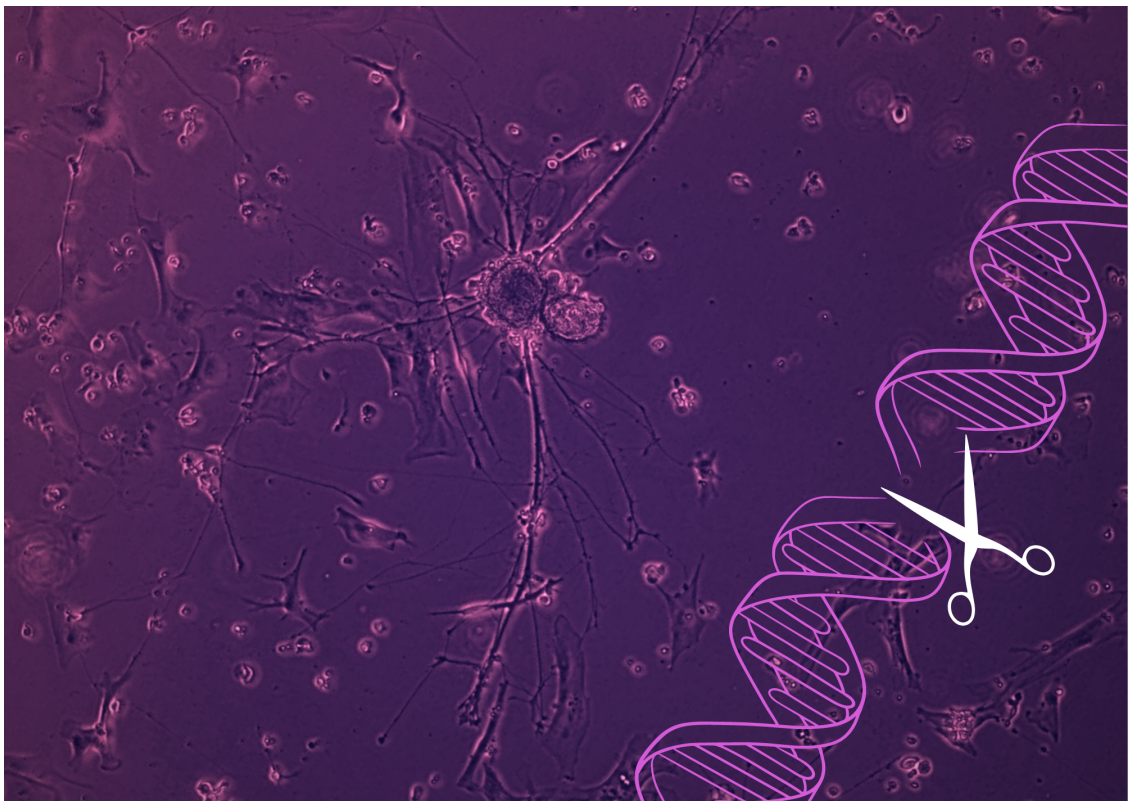
CHALMERS
UNIVERSITY OF TECHNOLOGY



Towards Safer Gene Therapies: Investigating Human Cell Type Effects on CRISPR-Cas9 Off-Target Editing

Master's thesis in Biotechnology

JULIA STEVRELL

DEPARTMENT OF LIFE SCIENCES

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2023

Master's thesis 2023

Towards Safer Gene Therapies:

Investigating Human Cell Type Effects on CRISPR-Cas9 Off-Target Editing

JULIA STEVRELL

CHALMERS
UNIVERSITY OF TECHNOLOGY

MSc Thesis 2023
Department of Life Sciences
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone +46 (0)31-772 1000

Sponsored and performed in collaboration with the Therapeutic Editing Team
Genome Engineering Department
AstraZeneca AB R&D
SE-431 50 Mölndal
Sweden
Telephone +46 (0)31-776 1000

Towards Safer Gene Therapies: Investigating Human Cell Type Effects on CRISPR-Cas9 Off-Target Editing
JULIA STEVRELL

Supervisor: Alexandra Madsen, Post-Doctoral Fellow, AstraZeneca, Gothenburg
Examiner: Yvonne Nygård, Associate Professor, Department of Life Sciences, Chalmers

Cover: Microscopy image of human iPSC-derived midbrain dopaminergic neurons. The colors have been altered for artistic effect.

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2023

Towards Safer Gene Therapies:
Investigating Human Cell Type Effects on CRISPR-Cas9 Off-Target Editing

JULIA STEVRELL
Department of Life Sciences
Chalmers University of Technology

Abstract

The CRISPR-Cas9 technology for precise and efficient genome editing holds immense potential in a vast range of bioscience applications, including the therapeutic treatment of somatic genetic disorders by directly correcting disease-causing mutations. However, the potential for CRISPR-Cas9-based gene therapies is largely limited by the risk of introducing unintended mutations at off-target genomic sites. These so-called *off-target effects (OTE)*, and their downstream phenotypic implications are key safety concerns for prospective therapeutic applications. Despite extensive research efforts, many potential underlying factors behind CRISPR-Cas9-mediated off-target activity in humans remain unexplored. To the best of our knowledge, cellular heterogeneity is one such incompletely evaluated factor. As such, this thesis project aimed to investigate the potential influence of different cell types on the off-target activity of both target-specific and unspecific ('promiscuous') CRISPR-SpCas9 genome editors. To this end, three different human cell lineages; ventricular cardiomyocytes, hepatocytes, and midbrain dopaminergic neurons, were generated from previously established engineered human induced pluripotent stem cell (hiPSC) lines through directed differentiation. hiPSC line generation followed a dual approach with TET-ON controlled expression of SpCas9 and U6-controlled expression of cell line-specific sgRNAs (PCSK9, VEGFA2, or HEK4) for transient, doxycycline-inducible genome editing in mature cell types. Differentiated cell types were partly validated with comparative gene expression and protein expression assays and subsequently assessed for on- and off-target editing with short-read, targeted Next-Generation Sequencing methods. On-target editing was validated with Amplicon-seq and showed significant cell type- and batch-dependent induction efficiencies as well as considerable leakage-mediated editing across cell types. Guide-specific off-target cleavage sites were identified with CHANGE-seq and cross-validated with independent methods. Small indel and SNP mutations at on- and off-target loci were verified with the rhAmpSeq assay and resulting OTE profiles were compared between cell types. Several significant cell type-specific off-target edits were detected for both target-specific and unspecific CRISPR-SpCas9. Although further validation is needed, the identified cell type-specific edits suggest that the cell type has an influence on CRISPR-Cas9 off-target nucleolytic activity. These findings could have considerable implications for continued research into the safety of CRISPR-Cas9 genome editors, and nevertheless their clinical use for combating genetic diseases.

Keywords: CRISPR-Cas9, hiPSC, differentiation, doxycycline, off-target editing, CHANGE-seq, rhAmpSeq, next-generation sequencing.

Acknowledgements

I express my sincere gratitude to my supervisor, Alexandra Madsen, for her invaluable guidance and wisdom throughout this project. Her support, kindness, and willingness to entrust me with responsibility have made this year the most enriching and educational experience of my academic journey so far.

I extend my thanks to the entire AstraZeneca team, particularly the BioPharmaceuticals R&D area, for providing me with the opportunity to engage in hands-on research and gain knowledge and experience within the pharmaceutical industry at an early stage in my career.

I would like to acknowledge and appreciate the invaluable advice and encouragement provided by the members of the Genome Engineering Department. I am especially grateful to Pinar Akçakaya, Alexandra Madsen, Niklas Selfjord, Marta Martinez-Lage Garcia, Yousra Yahia, Silvia Pedro, Davide Giorgio Berta, Nina Akrap, Sandra Wimberger, Pei-Pei Hsieh, Martin Peterka, Chinnu Rose Joseph, Arun Kumar Sundraramurthy, and Sasa Svikovic from the Therapeutic Editing Team. Their continuous feedback, creative ideas for further assays, and expertise in troubleshooting and validating experimental results have greatly contributed to the success of this project.

I extend my appreciation to the NGS team in the Translational Genomics Department, including Julia Lindgren, Julia Liz Touza, Margherita Francescatto, Inken Dillmann, Sanuja Ahammu, Marta Potapo, and Daniel Jachimowicz. Their coordination of bi-weekly sequencing work and assistance with the final rhAmpSeq assay have played a significant role in the project's outcomes.

I would like to thank Mike Firth from the Quantitative Biology Department for his expertise in bioinformatics analysis of the sequencing experiments conducted throughout this project.

I express my gratitude to my examiner, Yvonne Nygård at Chalmers, for her excellent support and mentorship in both the academic and administrative aspects of this project. Her guidance has been instrumental in preparing me for my future endeavors as a research scientist and bioengineer.

Lastly, I extend my deepest thanks to my family and friends for their unwavering support throughout this journey. I am especially grateful for their genuine interest and involvement in the health and well-being of my precious hiPSCs.

Julia Stevrell, Gothenburg, June 2023

List of Acronyms

Below is a list of acronyms used throughout this thesis:

Ab	Antibody
bp	Base-pair
Cas	CRISPR associated protein
CDR	Cardiomyocyte dissociation reagent
CHANGE-seq	Circularization for high-throughput analysis of nuclease genome-wide effects by sequencing
CIRCLE-seq	Circularization for in vitro reporting of cleavage effects by sequencing
CM	Cardiomyocyte (here hiPSC-derived ventricular cardiomyocyte)
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
crRNA	CRISPR RNA
DNA	Deoxyribonucleic acid
Dox	Doxycycline
DSB	Double-strand DNA break
GCDR	Gentle Cell Dissociation Reagent
GE	Genome Engineering
GUIDE-seq	Genome-wide, unbiased identification of DSBs enabled by sequencing
HDR	Homology Directed Repair
hiPSC	Human induced pluripotent stem cell
Hep	Hepatocyte (here hiPSC-derived)
indel	Insertions and deletions
NGS	Next Generation Sequencing
NHEJ	Non-homologous end joining
Nr/NR	Neuron (here hiPSC-derived midbrain dopaminergic neuron)
OTE	Off-target effects (referring here to unintended CRISPR-Cas9-mediated genomic mutations)
PAM	Protospacer adjacent motif
PCR	Polymerase chain reaction
PCSK9	Proprotein convertase subtilisin/kexin type 9
qPCR	Quantitative PCR
RGN	RNA-guided nuclease
RNA	Ribonucleic acid
sgRNA	Single guide RNA
SNP	Single nucleotide polymorphism
SpCas9	*Streptococcus pyogenes* CRISPR associated protein 9
SNV	Single nucleotide variant
TGE	Therapeutic Genome Editing
TIDE	Tracking of Indels by DEcomposition
tracrRNA	Trans-activating CRISPR RNA
VEGFA2	Vascular endothelial growth factor A, site 2

Contents

Keywords: CRISPR-Cas9, gene editing, off-target effects (OTE), hiPSC, *in vivo*, *ex vivo*, *in vitro*, NGS.

1
Introduction

In this chapter, a brief introduction to this thesis project is presented, including a background to the CRISPR-Cas9 system and its potential use- and current limitations for employment within the area of human therapeutic genome editing. Moreover, the importance of investigating cell type effects on CRISPR-Cas9 genome editing outcomes is introduced. Finally, the project aim-, research objectives- and study limitations are stated.

1.1 Project background

The discovery of the bacterial-derived CRISPR-Cas9 system and its application as a novel genome editing tool has been met with great enthusiasm and high expectations for possible therapeutic applications in a wide range of diseases [23]. This two-component system, consisting of a single guide RNA and a Cas9 nuclease, allows for targeted editing of DNA sequences of interest by introducing DNA double-strand breaks at the target site and is therefore generally referred to as CRISPR RNA-guided nucleases (CRISPR RGNs). Subsequent repair by either error-prone or precise cellular DNA repair mechanisms then facilitates the disruption of normal gene function or introduction of novel genetic features, thus potentially altering the cellular phenotype [54].

Due to their programmable nature and ease of use, CRISPR RGNs have a seemingly endless application spectrum, much like a molecular swiss-army knife. Not the least within medicine and drug discovery, where for example, the CRISPR-Cas9 system can be used to generate novel and more accurate models of disease pathophysiology [24]. These models provide novel insights into diseases and can also be used to screen new drug targets before entering clinical trials. However, the ultimate goal at AstraZeneca and other pharmaceutical companies is therapeutic editing; to correct genetic mutations and thus treat the very source of congenital- and acquired genetic diseases. AstraZeneca, and specifically the *Therapeutic Editing team* within the Genome Engineering Department, is currently exploring the possibilities of using the CRISPR-Cas system to advance gene therapies *in vivo* in humans, to cure, rather than treat, patients with genetic diseases [101].

Research into applying CRISPR-Cas9 as a therapeutic tool for curing somatic genetic diseases has progressed rapidly in the past decade. Yet, CRISPR-Cas9-based gene therapies are in the early stages of drug development, with most projects still in target discovery/target validation and proof of concept/proof of principal phases, due to safety and efficacy concerns. As with every new technology, there are critical

aspects that must be considered through rigorous research, before moving on to treating human patients. One of the most prominent concerns revolves around the safety aspects of therapeutic editing *in vivo* [18].

Avoiding changes occurring in the genome other than those intended, and thus running the risk of introducing new harmful mutations, is a technical difficulty associated with editing genes using the CRISPR technology. Unintended changes include 'wrong' changes at the 'right' location in the genome (on target) and changes elsewhere in the genome other than at the intended location (off target) [18]. The risk of undesirable changes is particularly great if several genes are edited simultaneously, and furthermore, if the edits are heritable, that is, editing germ cells (sperm and/or oocytes). Thus, most CRISPR-based therapeutic applications so far focus on monogenic and endogenous (hereditary or chronic) diseases and infectious diseases, with somatic editing strategies being the only ethically accepted approach (non-heritable genetic interventions). This includes therapies against beta-thalassemia and sickle-cell disease [27], cystic fibrosis [36], Huntington's disease [93], Duchenne's Muscular Dystrophy [80], transthyretin amyloidosis [33], congenital eye disease[47] and proviral infections like HIV [46]. Of these, therapeutic candidates for beta-thalassemia, sickle-cell disease, congenital eye disease, TTR, and HIV have already reached clinical trials. However, when performing targeted editing *in vivo*, there is a risk of introducing unintentional- and potentially harmful mutations, such as indels and substitutions, or larger edits, such as chromosomal rearrangements, at both the intended (on-target) site, and at unintended sites in the genome. The latter type of edits are collectively referred to as off-target effects (OTEs) and these effects could result in cells with reduced fitness, oncogenesis, or loss of function [18]. Therefore, off-target effects with CRISPR-based genome editing tools designed for therapeutic applications must be minimized and preferably eliminated to ensure the safety of treated patients.

It has been extensively demonstrated that CRISPR-Cas systems can introduce off-target effects with varying frequency, for example, [2, 16, 19, 29, 45, 71, 113] . Yet, as of today the underlying mechanisms behind such effects are still insufficiently understood. There is also a technical barrier to accurately quantifying less frequent off-target effects and that is the inherent error rate of most Next-Generation Sequencing tools (0.1% for indels measured with conventional Illumina®platform). This inevitably reduces the fidelity in measurements of lowly abundant OTEs. Thus, more sensitive off-target detection assays are needed to validate previously found, lowly abundant OTEs. What is also missing is an understanding of the genetic mechanisms and phenotypic consequences of off-target effects in clinically relevant tissues and organisms (humans). Previous reports investigating OTEs in different cell lines and non-human organisms have shown or assumed that biological replicates, that are consisting of many different cell types, exhibit the same off-target editing profiles. Yet, the human body consists of roughly 40 trillion cells [10] and around 200 different cell types[40], each with distinct characteristics; morphology, cellular stage, gene expression profile, chromatin structure, epigenetic modifications, metabolic fingerprint etc. [67, 40]. Despite this diversity, there is a lack of research into the influence of the cell type on CRISPR-Cas9 editing specificity- and efficiency. That is, to the best of our knowledge, it is not known whether the same CRISPR-

Cas9 on- and off-target editing profiles are observed in different human cell types. For example, it can be hypothesized that OTE as a result of erroneous homology-directed repair of CRISPR-Cas9-mediated DSBs will be more prevalent in mitotic cells, as opposed to quiescent cells. Furthermore, low-frequency off-target effects might have previously been missed simply because they occurred rarely in certain cell populations that have not yet been studied in the context of CRISPR-Cas9 off-target editing.

The aforementioned knowledge gap within CRISPR-Cas9-mediated off-target activity is of high clinical relevance for the long-term goal of developing CRISPR-based gene therapies for human genetic disorders. With in-depth knowledge of all potential mechanisms influencing CRISPR-Cas9 mediated off-target editing, one can improve the design of both guide RNAs and Cas-variants, as well as delivery strategy, for safer and more efficient therapeutic genome editing. For example, optimize the *in vivo/ex vivo* CRISPR-delivery strategy to be more cell- or tissue specific. Such strategies are of key concern to the pharmaceutical industry in general, and to the Therapeutic Editing team at AstraZeneca in particular, whose purpose is to develop novel gene editing approaches & technologies for therapeutic use.

1.2 Aim

In light of this background, the aim of this MSc Thesis project was to expand on our understanding of the mechanisms behind CRISPR-Cas9 off-target editing, as a prerequisite for developing safer human gene therapies. Specifically, the research purpose was to investigate the potential influence of cell type on CRISPR-SpCas9-mediated off-target effects in human iPSC-derived cell lineages. Our hypothesis was that the cell type has a significant impact on CRISPR-Cas9-mediated off-target nucleolytic activity in human cells, especially given that different cell types have unique chromatin structures and epigenetic profiles. Would our hypothesis be proven correct, or at least an insinuation of cell-type-specific off-target activity be observed, such findings could have a noteworthy implication for continued research about CRISPR-Cas9 specificity and safety, and nevertheless affect clinical development of CRISPR-Cas9 as a therapeutic tool to combat disease.

1.3 Objectives

The overall objective of this experimental MSc Thesis Project was to establish and validate an *in vitro*-based experimental pipeline for investigating potential cell type effects on CRISPR-Cas9 off-target activity in clinically relevant cell types. To achieve this overall objective, and stated research aim, five interim objectives were defined:

1. To assess the trilineage differentiation potential (pluripotency) of three established AZ-proprietary hiPSC lines, by generating cell types within the three different germ cell layers; endoderm, mesoderm, and ectoderm.
2. To establish three monolayer cell type-directed differentiation protocols for the same ODIn-hiPSCs, towards hepatocytes, ventricular cardiomyocytes, and

midbrain dopaminergic neurons.
3. To identify and validate 100-150 potential off-target cleavage sites for one specific and two unspecific/'promiscuous' CRISPR guide RNAs using the 'circularization for high-throughput analysis of nuclease genome-wide effects by sequencing' (CHANGE-seq) *in vitro* assay.
4. To use dox-inducible CRISPR-SpCas9 gene editing technology to introduce on/off-target editing events in the differentiated cell types and verify on-target editing with targeted deep sequencing.
5. To apply the IDT-proprietary RNAse H2 Amplicon Sequencing (rhAmpSeq) technology to investigate potential cell type effects on CRISPR-SpCas9 off-target activity in clinically relevant human cell types.

1.4 Author contribution

The project was carried out within the Genome Engineering Department at AstraZeneca, within the group of Therapeutic Editing. This team is led by Director Pinar Akcakaya and the project was supervised by Post-Doctoral Fellow Alexandra Madsen, Ph.D.

Cell-line (hiPSC) generation and on-target editing validation in generated ODIn-hiPSCs was performed by A. Madsen before the author's engagement with this project, in collaboration with colleagues in the Therapeutic Editing and Gene Editing Technologies team [64, 66].

ODIn-hiPSC pluripotency validation and cell-type directed differentiation was performed by the author, J. Stevrell. Cell characterization of differentiated cell types and library preparation for downstream on/off-target editing validation was also performed by J. Stevrell.

The wet-lab part of the CHANGE-seq assay for guide-specific off-target site identification was performed by N. Selfjord (TE team) and Julia Liz Touza (NGS-team). The bioinformatics analysis was performed by Margerita Francescatto (NGS team) and the interpretation of results, followed by rhAmpSeq panel design was performed by J. Stevrell (PCSK9 guide) and A. Madsen (VEGFA2 and HEK4 guides).

All targeted sequencing work was performed by the NGS team at the Translational Genomics Department, AstraZeneca, Gothenburg. NGS libraries for on/off-target AmpSeq/rhAmpSeq were prepared by the author and then forwarded to the NGS team for sequencing.

Finally, primary sequencing data for the AmpSeq and rhAmpSeq assays were analyzed through an AZ-proprietary bioinformatics pipeline by M. Firth from the Quantitative Biology Department (AZ). Subsequently, the preprocessed data were further analyzed (filtering, normalization, variant calling, and graphical visualization of mutation frequencies etc.) by the author on the same AZ proprietary platform.

2
Theory

In this chapter, a brief theoretical background to the principal concepts and methods underpinning this study is presented, to facilitate the understanding of the experimental results. Considering the project's aim and limitations, the emphasis is on reviewing the current understanding of unintended CRISPR-Cas9 human genome editing, and its implications for therapeutic applications. Furthermore, the known mechanisms behind CRISPR-Cas9 off-target effects and established methods for detecting them are reviewed. Finally, a theoretical foundation for the experimental assays applied to investigate potential cell-type effects on CRISPR-Cas9 off-target editing in human iPSC-derived cell lineages is provided.

2.1 Theoretical background

2.1.1 CRISPR-Cas9

Like most great inventions, the discovery of CRISPR occurred almost simultaneously and in three independent laboratories around the globe between 1990 [41, 51, 78]. Yet, the term "CRISPR" was not coined until 2002, and stands for *Clustered Regularly Interspaced Palindromic Repeats*. CRISPR refers to the repetitive DNA sequences (spacers) of the DNA region adjacent to the *cas* genes (CRISPR-associated genes) in the CRISPR locus [52, 75]. CRISPR-Cas was suggested to play a role in prokaryotic immunity against viruses in 2006 [70], and this hypothesis was proved two years later [12, 72]. The CRISPR prokaryotic defense mechanism was demonstrated to work by accumulating spacer sequences from invading phages or plasmid DNA, forming "molecular memories" for combating future infections. These spacers are transcribed into CRISPR RNAs (crRNAs), which guide Cas proteins to the target sequence and cleave the foreign DNA to interrupt viral infection [12, 72]. Garneau et al. (2010) further demonstrated that double-stranded breaks (DSBs) can be generated three nucleotides upstream of the Cas-recognition site, namely the protospacer-adjacent motif (PAM) [31]. In 2011, the CRISPR-Cas9 system was found to contain another small RNA, trans-activating CRISPR RNA (tracrRNA), that forms a duplex with crRNA and guides Cas9 to the PAM [20]. Finally, in 2012, Emanuelle Charpentier and Jennifer Doudna assembled the CRISPR-puzzle and showed that the CRISPR-Cas system could be used as a genome editing tool [54], which would earn them the 2020 Nobel Prize in Chemistry. However, it was the Zhang lab that in the following year published the first method for using CRISPR-Cas as a tool for genome editing in mice and humans [17].

The Cas-nuclease is in nature guided by two separate RNA molecules, sometimes called dual-guide RNA (dgRNA). The dgRNA consists of a CRISPR RNA (crRNA) molecule, that specifies where the Cas-nuclease will cut a DNA sequence, and the trans-activating CRISPR RNA (tracrRNA) molecule, which attaches to the crRNA like a handle for the Cas-nuclease [54]. Since Doudna' and Charpentier's revolutionary publication in 2012, researchers have found a way to simplify this system by combining the crRNA and the tracrRNA to form one, continuous molecule, known as a single-guide RNA (sgRNA). As a result, genome editors only need one RNA to edit a particular DNA sequence [54]. As for the Cas-nuclease, evolution has led to a plethora of different Cas-variants, each with a different structure, modularity, and nuclease activity [69]. The most well-studied system to date, and in many applications the most efficient, is undoubtedly CRISPR-Cas9. Cas9 is a single, two-lobed protein made of several dynamic regions called domains: the recognition lobe (abbreviated REC) and the nuclease lobe (NUC). While the REC lobe is important for binding the guide RNA, the NUC lobe contains two separate nuclease domains called HNH and RuvC that form the two blades of the Cas9 "molecular scissor" [55].

In nature, prokaryotic cells will use whichever CRISPR-Cas system they possess to fend off invading phages and foreign plasmids. But in the lab, researchers are not limited to using just one system, and many successful attempts have been made to engineer natural Cas-nucleases and guide RNAs to create even more efficient genome editors [15]. Genome editing involves modifying a specific DNA sequence within a living cell, and various tools have been developed for this purpose. For example, zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) were considered effective genome editors before CRISPR [23]. However, ZFNs and TALENs require custom protein engineering for each target site, whereas CRISPR-based approaches only require modification of the guide RNA. Protein engineering is complex, while guide RNAs can be easily and inexpensively synthesized *in-house* or by third-party RNA synthesis companies. The efficiency, precision, simplicity, and flexibility of CRISPR-Cas tools have led to their rapid adoption by scientists worldwide, despite being only a decade since the first proof-of-principle study was published [23].

There are two parts to CRISPR-Cas-mediated genome engineering. First, creating a targeted cut in the genomic DNA, and second, utilizing the cell's natural DNA repair mechanisms to introduce a desired sequence or edit [17]. First, Cas9 (or another Cas enzyme) and its guide RNA are designed to be complementary to a certain target sequence of interest in the genome. The two customized parts are then delivered to the target cell either as one or separate DNA cassettes, as mRNA or as RNA and Cas-protein. Once inside the cell, the CRISPR-Cas system is expressed, modified, and folded into the final RNA-guided nuclease (RGN) complex that constitutes the genome editing tool. Subsequently, the nuclease-fused gRNA localizes and matches to an orthologous genomic sequence adjacent to the PAM site in a process called *PAM recognition*. The PAM is specific for each Cas-variant and must match the guide RNA perfectly to initiate cleavage. For Cas9 from the bacterium *Streptococcus pyogenes*, SpCas9, the PAM-sequence is 'NGG' [55]. Next, Cas9 unwinds the DNA helix at the site of PAM recognition and the guide RNA binds with the complementary DNA strand in a process called *base pairing*. Base

pairing starts from the region adjacent to the PAM (seed region) and upon complete binding, cleavage is initiated. If there is no PAM or insufficient complementarity between the gRNA and the adjacent DNA, Cas9 will usually continue searching elsewhere. However, Cas9 may occasionally tolerate a few mismatches, resulting in off-target activity. Cleavage at the target DNA strand is performed by the HNH domain, while the RuvC domain cleaves the non-target DNA strand. These cuts occur three nucleotides upstream (5') of the PAM and depending on the type of nuclease, this process may produce either blunt or "sticky" (overhanging) DNA ends [17, 71].

In mammalian genome editing, cellular repair proteins mend the double-stranded DNA break [17, 84]. This process is however imperfect and may result in alterations to the initial sequence, so-called edits, or mutations. There are several ways to mend a DSB, but the two predominant strategies are homology-directed repair (HDR) and non-homologous end-joining (NHEJ) [56]. NHEJ is the predominant repair pathway in most cell types. It involves directly ligating the broken DNA ends, often resulting in small insertions or deletions (indels), or substitutions (SNPs) at the break site. This repair method is efficient but random, making it suitable for disrupting genes. HDR on the other hand is restricted to mitotic cells but offers more precise repair. It relies on a repair template or another proximal genomic region with induced DSB and homology to the sequences flanking the DSB (overhangs). Cells utilize this homology to guide the repair process, resulting in seamless editings, such as large insertions, deletions, or structural rearrangements. HDR is useful for precise gene corrections and targeted insertions but may also result in deleterious large chromosomal alterations [56].

Of the approximately 20000 annotated genes in the human genome, mutations in over 3000 have to date been linked to disease. A particularly alluring application of CRISPR programmable nucleases is therefore the potential to directly correct harmful mutations in affected tissues and cells to treat- or even cure diseases with genetic predisposition [18]. The application of gene editing technologies for targeted modifications of genes in the context of medical treatments is collectively referred to as therapeutic genome editing, and the CRISPR-Cas9 toolbox has truly revolutionized this field. However, as mentioned in the *Introduction*, unspecific edits may result in adverse phenotypic effects and aggravated pathogenesis [18].

2.1.2 CRISPR-Cas9 off-target editing

When a programmed nuclease introduces an edit at an unwanted DNA site it is called an off-target effect or off-target activity ('OTE' for short). CRISPR-Cas off-target events occur due to sequence homology between snippets of the guide RNA (gRNA) and genomic regions other than the intended target protospacer sequence. Early studies into CRISPR-Cas9 off-target activity showed that Cas9 cleavage can cause mutagenesis at mismatched sites at similar or even higher frequencies than the intended target [29, 45]. It has also been shown that off-target regions contain up to six mismatches compared to the on-target sites, although off-target sites with fewer mismatches have a higher likelihood of binding and cleavage [102]. Various approaches have been developed- and are under development for identifying and

characterizing potential off-target effects for specific guide RNA sequences. Initial investigations into the off-target effects of Cas9 focused on human cancer cell lines, where the frequency of such effects was notably high due to impaired DNA repair pathways in tumor cells [107]. These studies showed that the specificity of CRISPR-Cas9 primarily depends on the seed sequence of the gRNA, located within 10-12 base pairs directly upstream of the PAM. This has been further verified with crystal structure studies and single-molecule DNA experiments, suggesting that while the PAM site is essential for the initiation of Cas9 binding, the seed sequence corresponding to the 3'-end of the crRNA complementary recognition sequence, directly adjacent to PAM, is also critical for subsequent Cas9 binding and activation of nuclease activity [113][106].

Although the targeting specificity of Cas9 is believed to be tightly controlled by the seed region of the guide RNA and the adjacent PAM, potential off-target cleavage activity could still occur on a DNA sequence with three to five base pair mismatches in the $5'-$distal part of the seed region[113]. Moreover, previous studies have demonstrated that different guide RNA structures can affect the likelihood of cleavage at on-target and off-target sites [16, 45]. Although significant improvements have been made with the SpCas9 protein and its variants, in terms of both editing efficiency and specificity, they have still been shown to bind and cleave DNA at off-target sites in the human genome [106].

As with on-target editing, an off-target event can be repaired via the NHEJ pathway, potentially resulting in a smaller mutation; indel, or substitution/SNP. Rather, if the edit occurs simultaneously with an on-target or a second off-target cleavage event, the off-target activity can generate a larger and more complex edit, such as chromosomal rearrangements, or a large deletion between the two cleavage sites [28]. Overall, the research on off-target effects can be divided into two separate focus areas: 1. evaluating the mechanisms behind off-target effects and 2. reducing off-target effects [18]. In this project, the focus was on the former.

Off-target effects could result in drastic adverse health effects on prospective patients. It can thus be argued that alongside the issue of efficiently delivering the CRISPR-Cas system to the target cells, off-target effects are the most pressing issues currently facing basic- and applied CRISPR-research fields, including therapeutic areas. This is for good reasons, if not evaluated thoroughly and mitigated, gene therapy with CRISPR-Cas9 might pose a great risk to the health and well-being of patients.

2.1.3 Off-target detection methods

Many approaches have been developed for the detection of genome-wide CRISPR-Cas9-induced off-target activity. To date, the predominant approach for identifying Cas9 nuclease off-target activity has been to computationally predict likely off-target sites based on sequence homology (*in silico prediction*), and then to assess any potential editing activity by enzymatic assays based on mismatch-sensitive endonucleases with Sanger sequencing, or targeted deep sequencing (*in vitro verification*) [29, 84]. These *a priori* predictions are critical as 98% of SpCas9 guide RNAs in human exons and promoters have at least one off-target site with three or fewer mismatches [11],

and previous Cas9 specificity studies collectively demonstrated that off-target sites with three or fewer mismatches are significantly more likely to be cleaved than more dissimilar sites [29, 45, 71]. Selecting the most unique target site possible is thus a valuable strategy for improving specificity since many gene editing applications have several possible guide RNAs capable of accomplishing the same experimental outcome.

The limitation of *in silico* prediction is that they only experimentally validate unintended mutations based on computed sequence homology, by algorithms that have been trained on limited data sets [113]. Furthermore, most computational off-target prediction tools do not adequately consider off-target sites with gaps, bulges, or alternative protospacer adjacent motif (PAM) sequences. Overall, they are biased and most often do not consider structural effects or the cellular context when predicting OT sites. An ideal off-target detection approach should be unbiased and examine the entire genome. In contrast, recent advances in cell-free *in vitro* approaches enable whole-genome screening for sites of DNA cleavage upon incubation with Cas9, guide RNA, and purified genomic DNA. Although highly sensitive, these methods do not account for cellular characteristics that present potential obstacles to accessing DNA such as chromatin and nuclear architecture. Thus, cell-free *in vitro* methods are mainly used as an initial identification step, to narrow down the potential off-target sites across the whole genome, prior to validating those OTEs with targeted deep sequencing. By contrast, *in vitro* cell-based methods deliver Cas9 and guide RNA to living cells to identify resulting off-target events in a particular cellular context. However, these approaches can have lower sensitivity and need additional components, which can be difficult to deliver to cells, and also limits this approach to studies with fewer samples [57]. Looking forward, assays relying on a dual approach by combining *in silico* and/or *in vitro* off-target site prediction with targeted sequencing for characterization of off-target effects in the desired cellular context, holds promise for becoming the standard due to their sensitivity and accuracy [104].

Such a dual strategy for identification and verification of off-target editing with CRISPR-Cas9 *in vivo* was described in a study by Akcakaya et al. (2018) [2]. The approach, 'verification of*in vivo* off-targets' (VIVO), involved the generation of a set of potential off-target sites *in vitro* using the off-target detection method CIRCLE-seq. Identified sites were subsequently evaluated *in vivo* by Illumina®-based targeted sequencing of genomic DNA extracted from livers of nuclease-treated mice. The strategy was used to assess off-target effects induced by RGNs with either a 'promiscuous' gRNA (gP), deliberately designed to have high off-target activity, as well as two 'specific' gRNAs designed to have higher orthogonality (gM and gMH) to a well-characterized target locus (mouse *Pcsk9*/ human *PCSK9*). The study found that by careful design of gRNAs, efficient *in vivo* on-target editing can be achieved with no detectable off-target mutations at 0.1% detection sensitivity [2].

2.1.4 CRISPR-Cas On/Off-Target Editing Mechanisms

Both bacterial immunity and human genome editing rely on the ability of the Cas protein to locate and cleave specific DNA targets guided by the sequence of its

RNA guide [45]. However, there are differences between natural prokaryotic defense systems and artificial genome editing systems in terms of delivery strategies, guide RNA composition, Cas protein variants, target sequence, and host cell characteristics such as cell type and stage [53]. The application of CRISPR systems for DNA editing in eukaryotic cells presents unique challenges that can influence the outcomes of editing. Notably, the presence of genomic DNA within the nucleus poses a significant hurdle for precise CRISPR delivery.

To address the challenge of intracellular delivery, various methods have been developed, which need to be optimized for the specific target cell, tissue, organism, and editing objective. Common approaches for delivering CRISPR-Cas systems to cells include plasmid transformation (for bacteria), electroporation (temporarily disrupting the cell membrane using an electric voltage), chemical delivery (e.g., encapsulation in lipid nanoparticles), transduction with viral vectors (for human cells), *Agrobacterium* transformation (for plants), and microinjection of eggs or embryos (for animal models like mice or zebrafish) [65]. To overcome the challenge of intranuclear delivery, the addition of a nuclear localization signal (NLS), a short tag to the Cas9 protein, is commonly employed. The NLS tricks the cell's natural intranuclear delivery system into transporting the enzyme across the nuclear envelope. Once inside the nucleus, Cas9 can search for its target within the DNA. However, the delivery step of CRISPR systems remains a bottleneck in genome editing applications [65].

Eukaryotes possess more complex genome organization than bacteria. Their DNA is wrapped around histones and undergoes various epigenetic modifications, such as methylation and acetylation. This results in variations in chromatin structure, with some regions containing loosely packed DNA (heterochromatin) and others containing tightly packed DNA (euchromatin) [39, 40]. Chromatin structure plays a crucial role in gene expression regulation but also restricts the accessibility of DNA to Cas9 and other CRISPR enzymes. As a consequence, certain genomic target sites can be more challenging to edit than others. Strategies to overcome DNA accessibility issues are being developed, including novel approaches such as CRISPR-Cas-based epigenome editing [98], although this issue remains largely unresolved at present.

Genetic and cellular diversity also impact CRISPR-Cas on/off-target activity. Each individual possesses a unique genome, and every tissue in the human body consists of millions of cells with distinct characteristics [67]. Moreover, cells can transition among different biological states during their life cycle. Although our understanding of genetic and cellular diversity is incomplete, recent studies have demonstrated the significant influence of human genetic diversity on both on-target and off-target editing using CRISPR-Cas [13, 63].

2.1.5 Promises and challenges with CRISPR-Cas9 in therapeutic applications

Despite the risk of unwanted editing effects, the promise of using CRISPR-Cas9 systems as a therapeutic genome editing tool for preventing, treating, or even curing genetic disorders, infections, and cancers is promising and holds the potential to

improve the quality of life and potentially also save millions of lives in the future [18, 23, 86]. Yet, there is a long way ahead and a plethora of issues to solve, besides CRISPR delivery and DNA accessibility, before large-scale clinical use can become a reality.

The specificity of CRISPR-Cas as a genome editing tool is one of the main safety concerns for clinical application [18, 102, 106]. Genetic modifications are permanent, and deleterious off-target mutations could create cells with oncogenic phenotypes, reduced fitness, or functional impairment. Furthermore, oncogenic mutations resulting from off-target editing may lead to the expansion of edited cells, and thus even low levels of off-target mutagenesis may have devastating consequences. Missense mutations and gene silencing might instead prevent cells from carrying out vital functions, for instance, they might prevent pancreatic beta cells from producing insulin. The outcomes of such events pose a serious concern for clinical applications, especially in the field of therapeutic gene editing. If not addressed and evaluated thoroughly, gene therapy with CRISPR-Cas9 might pose a great risk to the health and well-being of patients [18, 102, 106].

2.1.6 Experimental Models in CRISPR-Cas9 Genome Editing

During early human embryogenesis, totipotent embryonic stem cells are formed, possessing the capability to differentiate into any cell type in the human body. Gastrulation, a later stage, involves the transformation of a one-dimensional layer of epithelial cells (blastula) into a three-layered structure known as the endoderm, mesoderm, and ectoderm. Each germ layer corresponds to the development of specific primitive systems during organogenesis. The endoderm gives rise to the gastrointestinal tract, gut lining, liver, pancreas, portions of the lungs, and glandular tissues. The mesoderm contributes to the musculoskeletal system, including connective tissue, non-epidermal integumentary components, the circulatory system, kidney, and internal sex organs. The ectoderm, the outer layer, generates the external ectoderm (epidermis, hair, nails) and the neuroectoderm (neuronal crest and neuronal tube-brain and spinal cord), as well as the lens of the eyes and the inner ear [77].

Induced pluripotent stem cells (iPSCs) are reprogrammed from differentiated somatic cells, typically adult fibroblasts [4], back into a pluripotent state similar to embryonic stem cells (ESCs). iPSCs possess the ability to differentiate into any cell type of the three embryonic germ layers (endoderm, mesoderm, ectoderm) and can proliferate indefinitely in cell culture. Unlike ESCs, which raise ethical and availability concerns due to their origin from pre-gastrula embryos, iPSCs offer flexible and non-invasive cell sourcing, providing more freedom and a wider range for downstream research and clinical applications. In 2006, Yamanaka's lab achieved the first reprogramming of iPSCs from mouse embryonic and adult mice fibroblasts through retrovirus-mediated transfection and subsequent expression of four genes encoding transcription factors Oct3/4, Sox2, c-Myc, and Klf4 [99]. The lab also successfully reprogrammed the first human iPSCs (hiPSCs) in 2007, using the same transcription factors as in the mouse iPSCs [100], marking a groundbreaking accomplishment that

opened up new avenues in biomedical research and drug development. iPSCs have diverse applications, including disease modeling, regenerative medicine, drug development, drug toxicity screening [42, 49], and more recently, CRISPR-Cas genome editing applications [42, 96].

Despite the success of human induced pluripotent stem cell (iPSC) technology, several challenges persist. One commonly overlooked challenge is the variability among iPSC lines in their capacity to differentiate into functional cells of a specific lineage, which is influenced by genetic background and reprogramming history and remains unpredictable [42]. To overcome this challenge, the use of isogenic or clonal iPSCs has emerged. Isogenic iPSCs exhibit limited genomic variability, typically characterized by a few single-nucleotide polymorphisms (SNPs), in contrast to embryonic or adult human pluripotent stem cells (PSCs) that can have millions of naturally occurring SNPs per cell [42]. This feature is particularly advantageous in genome editing research as it enables the study of specific mutations of interest while eliminating background noise [42].

Another challenge is the limited survival of human PSCs (hPSCs) in single-cell culture, which is the preferred condition for *in vitro* hPSC assays. However, the development of the Rho-kinase inhibitor Y-27632, which suppresses anoikis during the disaggregation of hPSC colonies, has significantly improved single-cell survival and facilitated long-term culture of hiPSCs and their subsequent differentiation applications [110].

Previous approaches to genetically engineering stem cells relied on inefficient homologous recombination. Fortunately, the development of more site-specific nucleases, such as CRISPR-Cas, has largely resolved this issue. These site-specific nucleases have not only improved somatic cell reprogramming into iPSCs but also enhanced the efficiency of genome editing [42].

Furthermore, iPSCs derived from somatic cells, such as fibroblasts, may retain certain characteristics and epigenetic features inherited from their embryonic origin [44, 59]. This poses concerns for downstream differentiation applications. However, recent discoveries indicate that the impact of epigenetic effects on cellular state is minimal, and isogenic iPSCs derived from fibroblasts exhibit similar differentiation capabilities to embryonic stem cells [61].

2.2 Experimental theory

2.2.1 CRISPR-Cas genome editing strategy and hiPSC line generation

A commonly used method for delivery of Cas9 to cells is through transient transfection of plasmids encoding the Cas-nuclease and single/dual gRNAs. This method is dependent on efficient transfection, which can be difficult to achieve in many cell types due to the size of the Cas9 proteins [65]. An alternative method to Cas9 delivery is to generate stable cell lines with inducible expression of the Cas9 nuclease. Generation of cell lines with inducible expression of Cas9 could for example be preferred in experiments requiring higher sample throughput and when targeting cell types where CRISPR delivery is a bottleneck to efficient editing.

An approach to simplify the generation of isogeneic hiPSC lines with inducible Cas9 nucleolytic activity was established by González et al. in 2014 [34]. Here, a TET-ON-controlled Cas9 expression cassette was integrated into the *AAVS1* locus of hiPSCs. Cas9 was thereby transiently expressed in the presence of the tetracycline-class antibiotic doxycycline (dox) and subsequent nucleolytic activity induced by the co-expression or delivery of a sgRNA. This system relies on the integration of a TET-inducible promoter controlling the expression of either the guide RNA or the Cas nuclease if integrated into separate loci (dual approach), or both if integrated into the same locus (all-in-one construct). This dox-inducible approach was used to generate hiPSCs with transient CRISPR-SpCas9 editing in a study by Lundin et al. in 2020 [66]), which lay the foundation for this thesis project. The ODIn hiPSCs contain a modular vector with a TET-ON tetracycline-inducible promotor and SpCas9, flanked by genetic insulators. This modular vector was integrated into the *AAVS1* locus of the hiPSCs using *Obligate Ligation-Gated Recombination* (ObLiGaRe). TET-ON systems for controlled expression of synthetic genes have been extensively used in synthetic biology applications [34, 35, 66]. Moreover, the *AAVS1* locus for Spcas9-integration was chosen as it is regarded as the "safe harbor locus" for DNA integration in human genomes [94].

An approach to then select the cells with the inducible CRISPR-Cas construct or any trans gene for that matter, is to utilize negative selection methods such as toxin-mediated cell death. One such approach is the Xential-method[64]. The Xential method takes advantage of diphtheria toxin and the selective expression of the *HBEGF* gene to enable clonal selection of recombinant cells. The *HBEGF* gene encodes a heparin-binding EGF-like growth factor, which mediates the cellular response to diphtheria-toxin (DT). The wild-type phenotype is lethally sensitive to the DT-toxin while a knock-in mutant is resistant. By utilizing the toxic properties of diphtheria, this technique provides a means to eliminate non-recombinant cells and enrich for isogenic cells carrying the desired genetic modification, in this case containing a sgRNA cassette [64].

Together, the generation of stable and isogenic cell lines with inducible expression of Cas9 and continuous expression of sgRNA can be achieved with a two-stage recombination strategy. The benefits of using such an inducible CRISPR-Cas system are foremost that it allows for simultaneous temporal regulation of gene editing [66].

Furthermore, the use of established cell lines avoids the need for *de novo* genetic cloning/cell line generation, which is beneficial from a time-efficiency perspective.

2.2.2 Generation of hiPSC-derived cell lineages through differentiation

Stem cell differentiation is the process by which embryogenesis can be mimicked in controlled laboratory settings through either directed- or undirected differentiation. Directed differentiation methods that utilize knowledge of endogenous developmental pathways, but instead result in the development of a certain cell type exclusively. In theory, human iPSCs have the potential to differentiate into any type of cell in our body, and assays have been developed to differentiate iPSCs into specific cell lineages, including neurons, cardiomyocytes, chondrocytes, retinal pigment epithelial cells, pancreatic islet cells, and hepatocytes [91, 97]. Several approaches, including standard 2D culture and 3D embroid body formation, have been explored for optimal differentiation, but issues still remain.

2.2.2.1 Validating differentiation

To verify the ability of iPSCs to differentiate into different cell lineages, newly reprogrammed iPSC and cell lines created from them should be subjected to quality control assays [91]. Furthermore, iPSC-derived cell lineages are often characterized using similar assays to ensure their validity as models for primary cell types. An important validation strategy for iPSCs is measuring pluripotency by the presence and quantity of surface markers such as Tra1-60, Tra1-81, SSEA3, and SSEA4, and/or transcription factors such as Oct4, Sox2, Nanog, and Lin28a [79, 83]. Equivalently, differentiated cell types can be characterized using cell-type specific cell surface- or intracellular markers that are unique and stably expressed in the mature cell types, e.g. [7, 14, 90, 92]. By comparing the expression of cell-specific markers and pluripotency markers in both differentiated and undifferentiated cell types, the success of a differentiation assay can be validated. For successful differentiation, cell-specific markers should be highly expressed in differentiated cell types and pluripotency markers lowly expressed.

2.2.2.1.1 Relative gene expression analysis with qPCR Quantitative real-time PCR (qPCR) is a well-developed method for gene expression analysis, in either absolute terms (i.e., numbers of copies of a specific RNA per sample) or relative terms (i.e., sample 1 has twice as much mRNA of a specific gene as sample 2) [22]. By far, the majority of analyses use relative quantitation as this is easier to measure accurately and often the most relevant. This is the case for comparing gene expression levels between cell types, for example when validating an iPSC differentiation assay [109]. In qPCR, the transcripts of targeted reporter genes are first converted to cDNA by conventional reverse-transcription PCR. The synthesized cDNA fragments are then duplicated in a highly specific and sensitive PCR reaction, where the quantitation is carried out in real-time.

The most common method for relative quantitation of gene expression by qPCR

is the C_T-method [22]. This method relies on two key assumptions: First, the PCR reaction is occurring with 100% efficiency. That is, with each cycle of PCR, the number of products doubles. Second, there is a gene (or genes) that is expressed at a constant level between the samples (control vs. treated). This endogenous control will be used to correct for any difference in sample loading. The choice of endogenous control is important and is often one or a few "housekeeping" genes that are necessary for normal cell function, commonly *PGK1* and/or *GAPDH* [81]. A C_T-value is established for each qPCR reaction and this value can be used to generate a relative gene expression level between treated samples and the negative control with C_T-analysis. The C_T is defined as the number of cycles that it takes for each reaction to reach an arbitrary amount of fluorescence and the C_T-method enables relative quantitation of gene expression between treated- and control samples. That is, the C_T-method allows for the determination of the X-fold gene expression of gene 'i' in the treated sample of interest compared to a control sample. This is achieved by normalizing both conditions with the expression of the endogenous control 'j' and then normalizing the normalized expression of gene 'i' in treated samples against the equivalent expression in the control condition.

For the purpose of validating successful hiPSC differentiation, "treated sample" refers to the differentiated specimens, whereas "untreated sample" refers to undifferentiated hiPSC (control). Finally, to obtain the fold-change gene expression, the $\Delta\Delta C_T$-value is raised to $2^{-\Delta\Delta C_T}$. A positive fold-change can be interpreted as higher expression of the lineages-specific genetic marker and thus successful differentiation, when compared to the corresponding expression in iPSCs. Note that qPCR is only a semi-quantitative analysis method and not an absolute indication of gene expression levels in the treated cells. Note further that this method only considers population average expression levels and does not discern intra-sample variability. This implies that there is a risk of overestimating- (or underestimating) the success of differentiation based on qPCR results solely. As such, this assay can be complemented by other single-cell characterization assays such as immunocytochemical assays, like flow cytometry.

2.2.2.1.2 Immunostaining and flow cytometry analysis The expression levels of specific cell surface markers or intracellular proteins can be assessed both quantitatively and qualitatively, on a single cell level, using flow cytometry [1]. By staining differentiated cells with fluorescently-conjugated antibodies or dyes specific to a set of chosen differentiation markers and then quantifying their expression with flow cytometry, the proportion of differentiated cells out of the total cell population can be determined [60, 91]. Thus, flow cytometry provides valuable information about the extent and efficiency of the differentiation process and aids in the validation and characterization of differentiated cell populations.

2.2.3 Identification and validation of on/off-target editing

2.2.3.1 Identification of genomic off-target sites

In the *VIVO study* [2], the cell-free *in vitro* assay CIRCLE-seq was adopted for genome-wide screening of cleavage sites for both specific- and promiscuous guide RNAs targeting the mouse *Pcsk9*- and human *PCSK9* gene, respectively. Similarly, CHANGE-seq ('Circularization for High-throughput Analysis of Nuclease Genome-wide Effects by sequencing') [62] is a high-throughput cell-free *in vitro* method that detects CRISPR-Cas9 nucleolytic effects at both on- and off-target genomic sites. During CHANGE-seq library preparation, genomic DNA samples are fragmented and circularized. The circularized DNA is then linearized, amplified with PCR, and subject to library preparation by adding sequencing adapters. High-throughput sequencing then generates millions of reads, which are aligned to a reference genome. By comparing edited cells to a control sample, potential off-target cleavage sites with different DNA repair outcomes are identified. Further analysis, such as targeted deep sequencing or PCR followed by Sanger sequencing, can then determine the precise nature and frequency of the off-target events. CHANGE-seq exceeds CIRCLE-seq in terms of sequencing efficiency while maintaining advantages like DNA repair machinery-independent detection, simultaneous paired-end reading of the cleavage sites for reference-independent discovery, high sensitivity, and no requirements for costly DNA synthesis. These advancements make it valuable for identifying CRISPR-Cas9-based off-target effects [62].

2.2.3.2 Amplicon sequencing for on-target editing validation

Validation of genome editing is often performed using targeted, short-read Next-Generation Sequencing (NGS) methods, such as the Illumina®platform, or Sanger sequencing [43]. While such methods are capable of detecting small insertion and deletion (indel) events and single nucleotide polymorphisms (SNPs/substitutions), which are the most common outcomes of CRISPR-Cas9 genome editing, they may fail to detect larger genome aberrations [43]. Amplicon sequencing, AmpSeq for short, is a targeted short-read NGS method that uses PCR to create sequences of DNA called amplicons [48]. AmpSeq involves sequencing of amplified loci across samples of interest to enable the analysis of genetic variation in specific genomic regions. The method involves two PCR steps, the first PCR amplifies a specific locus of interest, e.g. *PCSK9*, while the second PCR indexes the amplified target for facilitating downstream sequencing analysis. The locus-specific primers contain $5'-$ overhang handles that are used to introduce sample-specific barcoded sequences during PCR amplification. Pooling of samples is possible as long as the target site sequence between amplicons is known and the sample-specific barcoded sequences allow individual samples in a pool to be identified after sequencing. This method is suitable when the locus of interest is known and a significantly variable genotype at the site is expected, i.e. on-target editing effects. Thus, this method is suitable for detecting and verifying CRISPR-Cas editing events at a target locus[95]. However, a significant limitation is the sequencing throughput, here meaning the number of samples that can be run in parallel, which is always single-plex (one locus/run).

Thus, for large sample sets and multiplexed amplicon sequencing, an alternative approach is more suitable, namely rhAmpSeq.

2.2.3.3 The rhAmpSeq method for sensitive and multiplexed CRISPR off-target editing analysis

The rhAmpSeq CRISPR Analysis System from Integrated DNA Technologies (IDT)[89, 50] is a proprietary technology that allows for quick and accurate quantification of CRISPR-Cas edits at both on- and off-target sites in genomic DNA. It utilizes the RNAse H2-dependent PCR chemistry (rhAmpSeq), which combines the advantages of amplicon-based enrichment and next-generation sequencing (NGS) to accurately detect and quantify CRISPR editing effects. Similar to conventional amplicon sequencing, rhAmpSeq generates amplicon libraries for targeted deep sequencing. However, compared to the AmpSeq method, which is always single-plex, rhAmpSeq allows for multiplexing. I.e. it is preferred when multiple loci ($\leq$ 500 off-target sites per indexed sample) are to be analyzed.

The rhAmpSeq method can be divided into four parts. First, rhAmpSeq panel design and synthesis. IDT offers both to design and/or synthesize rhAmpSeq panels designed by the customer. The rhAmpSeq panels consist of multiplexed PCR primers targeting both the genomic on-target- and potential off-target sites and are customized based on the specific genomic target and guide RNA sequences. Second, the preparation of sequencing libraries, where genomic DNA from the CRISPR-edited samples is extracted and used for library preparation. The rhAmpSeq primers are utilized to amplify the targeted regions of interest, including the on-target site and potential off-target sites.

Fundamentally, rhAmpSeq library preparation is a two-step PCR assay. In PCR 1, RNA-base-containing blocked primers (rhAmp primers) form duplexes with their genomic target site and are then recognized by the thermostable RNase H2 enzyme. Only perfectly matched primers-DNA duplexes are activated by RNase cleavage at the DNA:RNA duplex. The rhAmp-primers are then extended to create amplicons containing the target regions of interest. After cleanup, these amplicons are indexed and further amplified in PCR 2 to allow for the pooling of multiple samples. After another library cleanup, the pooled and indexed libraries are quantified and sequenced using Next-Generation Sequencing on the Illumina®NGS platform. This step generates sequencing data that includes information on the presence and abundance of on-target and off-target editing events, including read count and mutation variants (small edits). Overall, the rhAmpSeq CRISPR Analysis System from IDT offers a robust and user-friendly solution for evaluating off-target editing in CRISPR experiments. It combines rhAmpSeq primer design, library preparation, NGS, and data analysis to provide accurate and comprehensive insights into the specificity and off-target effects of CRISPR-Cas9 editing.

3
Methods

In this chapter, the materials- and methodological details are summarized. For a more detailed experimental disclosure, including a list of all reagents-, protocols, and genetic sequences used, as well as the safety-, ethical- and environmental disclosures, please consult the Appendices.

3.1 Cell line generation

Three different ODIn-hiPSC lines with dox-inducible CRISPR-SpCas9 systems were generated *in-house* as described in *Appendix A*, Section A.2 and provided by A. Madsen as frozen cryovials (−80°C). In brief, a recombinant SpCas9-cassette with the Tet-On 3G tetracycline-inducible expression system (Takara Bio) was integrated into the *AAVS1* locus of reprogrammed healthy human induced pluripotent stem cells using the ObLiGaRe-approach [66]. The chosen Cas-variant was Cas9 from the bacterium *Streptococcus pyogenes*, as it is the most widely adopted and well-characterized Cas9 homolog to date [15]. Henceforth, Cas9 and SpCas9 may be used interchangeably, note that it is the SpCas9-variant that is referred to. Next, the Xential method [64] was employed by A. Madsen and others to introduce either a target-specific or 'promiscuous' gRNA cassette into the *HBEGF* locus, under the control of a U6-promoter. This allowed for diphtheria-toxin-mediated negative selection of isogenic recombinant ODIn-hiPSCs.

The integrated sgRNAs were U6-PCSK9, U6-VEGFA2, and U6-HEK4, the former one designed to match a highly unique genomic sequence (no off-targets with less than 4 mismatches to the guide sequence) within the human gene *Proprotein convertase subtilisin/kexin type 9*, and it is an internal benchmark guide at AZ. The other two guides are externally validated benchmark guides for CRISPR-Cas9 off-target editing analysis studies [103, 105] and target the *Vascular Endothelial Growth Factor A, site 2* (*VEGFA2*), and the intergenic sequence HEK293-site4 (*HEK4*), respectively. U6-VEGFA2 and U6-HEK4 were intentionally designed to target unspecific genomic sites, that is, sites with high sequence homology to multiple genomic locations (1-3 mismatches). This unspecificity increases the probability of off-target binding and nucleolytic activity compared to specific guides like U6-PCSK9. The U6-VEGFA2 and U6-HEK4-guides are therefore referred to as 'promiscuous' while the U6-PCSK9-guide is referred to as 'specific'. Any eventual cell type-specific edits would likely be detected at least for the promiscuous guides. However, it should be noted that promiscuous guides would not be used in therapeutic applications.

3.2 hiPSC culture

The ODIn-hiPSCs were expanded prior to trilineage differentiation and cell type-directed differentiation for a minimum of two weeks in serum- and feeder-free conditions using the Cellartis®DEF-CS™500 Culture System (Cat. nr. Y30010; nr. Y30012; nr. Y30017, Takara Bio) following the manufacturer's instructions. In brief, cells were maintained as single-cell monolayer cultures on COAT-1 coated 6-well tissue-treated culture plates (Corning®Costar®). Daily media changes were performed with 0.4 $mL\,cm^{-2}$ of complete DEF-CS™media for maintenance, where DEF-CS™basal media had been supplemented with growth factors 1 (1:333) and growth factor 2 (1:1000) pre-warmed to 37°C. Cells were incubated in a humidified incubator supplied with 5% CO2. For passaging, cells were dissociated into a single cell suspension with TrypLE Select Enzyme (1X) (Gibco™) and resuspended in complete DEF-CS™media for thawing and passaging (growth factor 1 + 2 + 3, 1:333, 1:1000 and 1:1000). The Odin-hiPSCs were seeded at a density of $4 - 5 * 10^4$ $cellscm^{-2}$ in 3 mL DEF-CS™media for thawing and passaging. The cell culture medium was changed daily until the cells reached a confluence of $1.5 - 3.0 * 10^5$ $cellscm^{-2}$, which generally occurred 3-4 days after the previous passage.

3.3 hiPSC differentiation

3.3.1 Trilineage differentiation for pluripotency validation of the ODIn-hiPSCs

After cell expansion of the ODIn-hiPSCs for nine days, differentiation of the ODIn-hiPSCs into monolayer cultures of the three germ layers (endoderm, mesoderm, and ectoderm) was performed using the STEMdiff™Trilineage Differentiation Kit (Cat. nr. 05230, STEMCELL™Technologies). The purpose of this assay was to assess the pluripotency of the ODIn-hiPSCs as a quality control of their innate potential as iPSCs, before proceeding with the more lengthy directed differentiation towards cell types within the different germ layers (hepatocytes from the endodermal layer, ventricular cardiomyocytes from the mesodermal layer and midbrain neurons from the ectodermal layer). Note that the trilineage differentiation kit only provides a functional overall assessment of the pluripotency of the iPSCs and their ability to differentiate in the overall direction of the three germ layers, not necessarily ensuring a capability for further differentiation. The ODIn-hiPSCs were differentiated according to the manufacturer's instructions, for further experimental details see *Appendix A*, Section A.4. In brief, the ODIn-hiPSCs were plated onto Matrigel (Corning) with mTeSR1 on Day 0, and from Day 1 onwards fed with endoderm or mesoderm differentiation media once daily for 5 days, or ectoderm differentiation media for 7 days. Cells were then harvested with GCDR (Cat. nr. 100-0485, STEMCELL™Technologies) for cell characterization.

Derived germ lineages from all three cell lines were harvested on day 5 (endoderm and mesoderm cells) and on day 7 (ectoderm cells). Harvest was performed with the RLT lysis buffer (Qiagen) and RNA was extracted with the Qiagen AllPrep Mini kit

for comparative gene expression analysis with quantitative PCR (qPCR). The specific objective was to assess the relative expression of trilineage-specific marker genes in each cell line and cell lineage, respectively, and thus verify ODIn-hiPSC trilineage differentiation potential. For this purpose, 20 ng/reaction of mRNA extracted from the three ODIn-hiPSC-derived germ lineages (endoderm, mesoderm and ectoderm) and mRNA from ODIn-hiPSCs (provided by A. Madsen) were reverse-transcribed with the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems™). Probes (TaqMan™) were specific to seven different marker genes; *FOXA2*, *SOX17*, and *CXCR4* for the endoderm cells, *Brachyury (T)*, *CXCR4* and *NCAM* for the mesoderm cells, and *NCAM* and a *Nestin* for the ectoderm cells. Moreover, two probes targeting either of the two housekeeping marker genes; *PGK1* and *GAPDH*, were assayed for all cDNA samples for normalization purposes. In total, 3 biological replicates x 3 technical replicates x 3 cell lines x 3 cell lineages + 3 hiPSC control samples/probe + 3 water control samples/probe were pipetted on a 384-well qPCR plate (Cat. nr. AB1384, Thermo Scientific). Quantitative PCR was run on a QuantStudio 7 Flex machine according to the standard qPCR protocol, see *Appendix A*, Section A.4.3.1 for further details.

3.3.2 Cell type-directed differentiation

For the cell type-directed differentiation, commercial differentiation kits were used, see *Appendix A*, Section A.4.2, Table A.4 for a summary. All protocols were performed according to the manufacturer's instructions, with minor adjustments. For further experimental details, see *Appendix B*. A brief experimental summary is given below.

3.3.2.1 Cardiomyocyte differentiation

The ventricular cardiomyocytes were differentiated and maintained with the STEMdiff™Cardiomyocyte Differentiation Kit (Cat. nr. 05010 STEMCELL™Technologies), according to the manufacturer's instructions. This protocol offers a three-step differentiation procedure from hiPSCs to ventricular cardiomyocytes in 14 days. It is also optimized for use with the mTeSR™1/Matrigel hiPSC culture system. In brief, ODIn-hiPSCs were expanded in the Cellartis DEF-CS™500 System (Cat. nr. Y30020, Takara Bio) for 2-3 weeks. Two days prior to cardiac differentiation the hiPSCs were reseeded at a density of $4 * 10^5$ cells/well (diff 1), $3.5 * 10^5$ and $4.5 * 10^5$ cells/well (diff 2) onto Matrigel®coated 12-well flat-bottom TC-treated Corning®Costar®plates (one plate per cell line), and fed mTeSR™1 media supplemented with Y-27632. On Day 0, and every other two days for six days, full media changes (2 ml/well) proceeded with three different complete differentiation media, followed by maintenance media for at least an additional eight days. See *Appendix A*, Section A.4.2 for further details.

A second batch of ODIn-hiPSC-derived ventricular cardiomyocytes was generated with the same differentiation kit but with minor adjustments to the protocol. These changes included changing the initial hiPSC seeding density (two different seeding densities per cell line), maintaining the CMs for longer at the maturation stage, and

using the STEMdiff™Cardiomyocyte Dissociation Kit (STEMCELL™Technologies) instead of GCDR (STEMCELL™Technologies) for final cell-dissociation.

3.3.2.2 Hepatic differentiation

For hepatic differentiation, initially, the Cellartis®iPS Cell to Hepatocyte Differentiation System from Takara Bio, Europe (Cat. nr.Y30055, Takara Bio) was used. This is a 21-day, two-part protocol based on monolayer-, serum- and feeder-free culture conditions, optimized for use with the DEF-CS™™500 System for hiPSC culture. The kit was used according to the manufacturer's instructions. For further details see *Appendix A*, Section A.4.2. In brief, definitive endoderm (DE) cells were generated from hiPSCs expanded in the DEF-CS™500 System for 2-3 weeks. DE cells were obtained within one week, through daily media changes with definitive endoderm differentiation media (1, 2, 3, 4 and 6). On day 7, DE cells were passaged and hepatic differentiation proceeded for at least 14 days, with media changes performed every 2-3 days with three different supplemented hepatic differentiation media. Upon hepatic maturation, cells were grown as a monolayer under a sensitive gelatinous overlay.

A second hepatic differentiation was performed by A. Madsen due to the unfortunate disclosure of doxycycline-derivatives in the Cellartis®iPS Cell to Hepatocyte Differentiation System, which induced SpCas9 expression and nucleolytic activity during differentiation. The second hepatic differentiation assay was performed with a novel protocol, the STEMdiff™Hepatocyte Kit (Cat. nr.100-0520, STEMCELL™Technologies). This is also a 21-day and serum-free differentiation protocol for the generation of hepatocyte-like cells from human PSCs. Unlike the Cellartis®protocol, this is a three-step differentiation procedure, from hiPSCs to DE cells in 5 days, to hepatocyte progenitor cells (HPC) on day 10 and finally to mature hepatocyte-like cells (HLC) on day 21. In brief, the expanded hiPSCs were passaged onto Matrigel®-coated 24-well flat-bottom TC-treated Corning®Costar®plates (one plate per cell line) and fed mTeSR™1 media supplemented with Y-27632. Daily media changes were performed with different, supplemented hepatic differentiation media for 21 days (STEMCELL™Technologies).

3.3.2.3 Neuronal differentiation

Midbrain dopaminergic neurons were generated from ODIn-hiPSCs using the STEMdiff™Midbrain Neuron Differentiation System (STEMCELL™Technologies). This is a two-month, three-part protocol including the STEMdiff™SMADi Neuronal Induction Kit (Cat. nr. 05835, STEMCELL™Technologies), the STEMdiff™Midbrain Neuron Differentiation Kit (Cat. nr. 100-0038, STEMCELL™Technologies) and the STEMdiff™Midbrain Neuron Maturation Kit (Cat. nr. 100-0041, STEMCELL™Technologies). The monolayer differentiation approach was chosen due to time constraints and for better comparison with the other two cell types generated in the monolayer format. The differentiation protocols were performed according to the manufacturer's instructions (STEMCELL™Technologies) and repeated twice (from NPC stage/ differentiation stage) for optimization of the final cell density and subsequent RNA/DNA yield.

For both batches, initial seeding densities for neuronal induction of hiPSCs were $2 * 10^5$ cellscm^{-2} in a 6-well plate format with PLO/Laminin-coating (1 ml/well, respectively). For subsequent passages of the Neuronal progenitor cells, the plated cell density was $1.75 * 10^5$ cellscm^{-2}, also in 6-well plate format with PLO/Laminin coating (1 ml/well, respectively). For Neuronal differentiation, the first batch of neurons was seeded at $1.0 * 10^5$ cellscm^{-2}, while the second batch was seeded at a higher density of $1.25 * 10^5$. For Neuronal maturation, the first batch was seeded at $6 * 10^4$ cellscm^{-2} and the second at $8 * 10^4$ cellscm^{-2}, both diff. and maturation was carried out on PLO/Laminin-coated 6-well plates (1.5 ml/well, respectively). For further details, see *Appendix A*, Section A.4.2.

3.4 Induction of CRISPR-SpCas9 genome editing

After the generation of mature cell types, the objective was to induce SpCas9 expression and thus gene editing, with doxycycline. To get a sense of the tolerated dose of doxycycline, before cytotoxic responses are triggered, a dose-response assay was performed on the trilineage differentiated cells as well as a rigorous literature review on previous studies on dox-inducible gene editing in hiPSC-derived cell types, with a focus on hiPSC-derived cardiomyocytes, hepatocytes, and neurons.

3.4.1 Dox-titration for dose-response evaluation in trilineage differentiated ODIn cells

After the trilineage differentiation assay, a dox dose-response evaluation was performed on the remaining ODIn-PCSK9 cells (3x3 wells, 12-well plate format, Corning®) at day 5 of differentiation for the endodermal and mesodermal cells, and at day 7 for the ectodermal cells. The purpose was to find the highest tolerated dose for doxycycline induction of the final cell types, prior to the directed differentiation assays.

On the day of harvest for the trilineage differentiated cells, the old medium was aspirated and exchanged for 1 ml/well of dox-spiked STEMdiff™differentiation media, at four different concentrations (one biological replicate each); 1, 10 and 100 ng/mL for endodermal and mesodermal cells (day 5) and 0.1, 1 and 10 ng/mL for ectodermal cells (day 7). After 48h of incubation at 37°C, cells were harvested and genomic DNA simultaneously extracted with 100 µL/well QuickExtract (Lucigen™) and incubated at 37°C for 10 min. After incubation, cells were transferred to 1.5 mL Eppendorf tubes and vortexed for 15s. Each sample of genomic DNA was transferred to a PCR tube and incubated in a thermocycler at 70°C for 10 minutes, followed by 98°C for 10 min. The reaction was finally cooled down to 4°C.

The nine harvested ODIn-PCSK9 genomic DNA samples were provided with a unique combination of barcoded forward- and reverse-primers for AmpSeq and amplified with Phusion Hot Start II High-Fidelity PCR Master Mix (Cat. nr. F565S, ThermoScientific™). Sequencing libraries for targeted deep sequencing of the *PCSK9* locus, respectively, were prepared according to the internally-established NGS library preparation protocol, see Section A.5 for more details. Furthermore, a com-

plete list of primers, barcodes, adapters, and their sequences is provided for each target locus in *Appendix B*, Tables B.2 for *PCSK9*, B.3 for *VEGFA2*, and B.4 for *HEK4*, respectively.

3.5 Evaluation of cell type-directed differentiation

3.5.1 Cell morphology

All ODIn cell types were cultured in the same format; monolayer cultures were maintained in Corning®Costar®flat-bottom TC-treated culture plates. Morphological characteristics of the differentiated cell lineages were observed daily and photodocumented at iPSC-stage, progenitor cell stage (definitive endoderm and Neuronal progenitor cell stage), and mature cell stages (ventricular cardiomyocytes, hepatocytes, and midbrain neurons), using the EVOS™M5000 Imaging System from Invitrogen™(ThermoFisher Scientific). Time-lapse videos were also taken of the beating cardiomyocytes at the maturation stage. The imaging system used for this purpose was also an EVOSTM M5000 from Invitrogen™(Cat. nr.AMF5000, ThermoFisher Scientific) with 10x magnification.

3.5.2 qPCR

In this study, we were interested in the relative gene expression between differentiated hiPSC-derived cells and undifferentiated ODIn-hiPSCs. By comparing the relative gene expression of both lineage-specific-, and pluripotency genetic markers, while also normalizing against the expression of housekeeping genes and equivalent expression levels in hiPSC control samples, one can evaluate whether the differentiation from hiPSC to different cell lineages was successful or not. The choice of qPCR assays and fluorescent reporter probes was based on recommendations from each respective differentiation protocol if provided. When such recommendations were absent, the probes were chosen with reference to previously published literature. A detailed protocol- and description of the TaqMan™probes used for qPCR with both the trilineage cells and downstream cell lineages is provided in *Appendix A*, Section A.4.3.1.

In brief, three biological replicates per cell line and cell lineage were harvested using the RLT protocol (Qiagen). Total mRNA content was extracted using Qiagen AllPrep Mini- or Micro Kits. The mRNA was eluted with nuclease-free water and the RNA concentration was measured with a NanoDrop One/One^C Microvolume UV-Vis Spectrophotometer (Thermo Scientific™). ODIn-hiPCS mRNA without sgRNA cassettes was provided by A. Madsen and used to normalize the expression of cell-specific marker genes. Reverse transcription of 20 ng total RNA per reaction was performed using the High-Capacity cDNA Reverse Transcription Kit with RNase Inhibitor (Applied™Biosystems). If necessary, mRNA was diluted in RNase-free water. The cDNA synthesis protocol included incubation at specific temperatures, see *Appendix A*. The cDNA was assayed with qPCR using cell lineage-specific TaqMan MGB-binding Fluorescent Probes (Applied™Biosystems) on a QuantStudio 7 Flex qPCR machine (Applied™Biosystems). Moreover, the POU5F1-probe was

chosen for assessing pluripotency in the downstream ODIn cell lineages for assessing differentiation outcomes.

Three technical replicates per sample were assayed in a 384-well format with 2-3 lineage-specific probes, two housekeeping genes (*PGK1* and *GAPDH*), and a pluripotency marker gene (*POU5F1*). Control ODIn-hiPSC mRNA was included in all runs. Each qPCR reaction included cDNA, TaqMan™Fast Advanced Master Mix (Cat. nr. 4444557), a lineage-specific probe, and nuclease-free water. The qPCR protocol included specific temperature and time conditions, see *Appendix A*. The C_T-values obtained from the qPCR were exported and transformed into relative gene expression values using comparative Ct-analysis with the $2^{-\Delta\Delta C_T}$ method. Here, the output C_T-values were normalized against housekeeping genes and corresponding hiPSC control values to obtain relative fold changes in gene expression.

3.5.3 Immunostaining and flow cytometry analysis

To further validate cell type-directed differentiation on a single-cell level, i.e. characterize the proportion of cells in each population expressing a cell type-specific antigen, the mature cell types were stained with cell type-specific fluorophore-conjugated monoclonal antibodies and analyzed with flow cytometry. In brief, cells were harvested at the maturation stage of differentiation with a suitable dissociation protocol, see *Appendix A*, Section A.1.2. Next, up to $1 * 10^6$ cells per staining were pelleted and stained according to the antibody manufacturer's instructions, either with an intracellular or a surface stain protocol. For more details on the staining procedure, see *Appendix A*, Section A.4.4.

Besides cell type-specific antibodies, each assay included an isotype control stain (IgG) for assessment of unspecific antigen binding. Also, if enough cells, a pluripotency stain (SSEA4) was also performed to check for the proportion of remaining hiPSCs/pluripotent progenitors in the differentiated population. Furthermore, a positive control cell line for each lineage-specific marker (and IgG) was tested, to validate the quality and cell type-specificity of each antibody. U2-OS was assayed for cTNT, HepG2 for ASGPR1, and SH-SY5Y for TUBB3.

The ODIn-hiPSCs were stained with the SSEA4-PE conjugated antibody, to validate pluripotency. For this purpose, the Abcam surface staining protocol was used. For the immunocytochemistry analysis of both batches of differentiated ventricular cardiomyocytes, the intracellular cardiac troponin T (cTnT) protein was targeted using a PE-labeled human monoclonal antibody (Miltenyi) as well as a human monoclonal PE-labeled IgG antibody for isotype control of unspecific antigen binding (Miltenyi). To this end, the intracellular staining protocol from Miltenyi (first batch of CMs) was applied, and the Abcam intracellular staining protocol for the second batch. Due to concerns about insufficient fixation and staining with the Miltenyi intracellular staining protocol, the cardiomyocytes were fixed and stained for an hour each, instead of 10 minutes as specified by the original protocol. For staining of the second CM batch, a third staining with the SSEA4-pluripotency marker was performed, to characterize the proportion of pluripotent cells remaining after differentiation. Also, a U2-OS cell line was tested for both cTnT and IgG as a positive control.

The Cellartis®hepatocytes were stained with ASGPR-PE, IgG-PE, and SSEA4-PE (Abcam) using the surface-staining protocol from Abcam, without alterations to the original protocol. The STEMdiff™hepatocytes were stained with Albumin-FITC, IgG-FITC (Abcam intracellular protocol) and SSEA4-PE (Abcam surface protocol). The neurons were stained with TUBB3-APC, IgG-APC (Abcam intracellular staining protocol) and SSEA4-PE (Abcam surface staining protocol). Also here, all washing steps were performed once and not twice, as recommended by the original protocol. All stainings were performed in protection from artificial light and after staining, cells were stored under aluminum foil, in the fridge, until flow cytometric acquisition.

Flow cytometry and data analysis were performed on the same day as staining, with the BD FACSymphony A1 Cell Analyzer (BD Biosciences). The subpopulation of cell lineage-specific positive cells was gated based on forward scatter (FSC) and side scatter (SSC), removing smaller particles and unstained cells (negative for the cell-specific marker). A second gating where FSC-area vs. FSC-height removed any duplets and larger cell clusters to allow for single-cell analysis. For more details on the flow cytometry analysis procedure, see *Appendix A*, Section A.4.4.2.

3.6 Dox-induction of ODIn cell types

3.6.1 Cardiomyocytes

For activating recombinant TET-ON systems in iPSC-derived cardiomyocytes, literature sources reported having used a doxycycline dose in a range of 100 ng/mL to 1 μg/mL for 48h [74, 73]. However, internal research groups in applied stem cell research at AstraZeneca recommended a more conservative approach of 100 ng/mL for 48h (same dox-regiment as for ODIn-hiPSCs). As a cytotoxic cellular response (extensive cell death) was observed during the dox-titration assay on trilineage differentiated mesodermal cells already at 100 ng/ml, it was decided to proceed with the internal recommendations. Half of the wells of ODIn-hiPSC-derived ventricular cardiomyocytes (6-12 biological replicates/ cell line, 2 different initial seeding densities) were induced with dox-spiked maturation media (STEMCELL™Technologies). For the first batch of cardiomyocytes, induction was performed on Day 16 of the differentiation protocol. For the second batch, induction was performed on Day 18 to allow further maturation. Both batches were induced with 100 ng/mL dox, based on recommendations from the AZ Applied Stem Cell Research Team (Discovery Sciences, AZ, Gothenburg). The cells were incubated at 37°C for 48h, followed by a regular, complete media change and continued maintenance for up to a week prior to cell harvest for downstream targeted deep sequencing analysis and investigation of cell type on/off-target editing effects.

3.6.2 Hepatocytes

Literature suggested inducing stem cell-derived hepatocytes with 250-500 ng/mL dox[8, 82]. As cytotoxic effects were visible also for the endodermal cells treated with 100 ng/mL dox during the trilineage assay, the lower recommended dose (250

ng/mL) was chosen for inducing gene editing in the ODIn-hiPSC-derived hepatocytes (both batches). Half of the wells per cell line were induced with dox-spiked hepatocyte maintenance media on Day 26 of the Cellartis®hepatic differentiation protocol and on Day 21 for the STEMdiff™hepatic differentiation protocol, at a concentration of 250 ng/ml. Cells were incubated at 37°C for 48h followed by a regular, complete media change and continued maintenance for up to a week prior to cell harvest for downstream targeted deep sequencing analysis and investigation of cell type on/off-target editing effects.

3.6.3 Neurons

For the dox-induction of neurons, literature references did not provide a consensus recommendation. Rather, the reported dose ranged from 100 ng/mL to 2 μg/mL [3, 92, 87]. Therefore, the first batch of ODIn neurons was dosed more conservatively (500 ng/ml) due to the same cytotoxicity concerns as for the cardiomyocytes and hepatocytes. For the second batch of ODIn-derived neurons, the dose was increased to 1 µg/mL, as targeted deep sequencing for on-target editing validation of the first batch revealed low editing efficiency. Dox-induction was performed by spiking the neuron maturation media (STEMdiff™Neuronal maturation kit) with dox and then performing a half media change according to protocol. Half of the wells per cell line were fed regular neuron maturation media and half were fed dox-spiked maturation media, reaching a final concentration of approximately 500 ng/ml (batch I) and 1 µg/mL (batch II), with possible variations due to suspected partial evaporation of the old media. Both batches were induced for 72h at 37°C, followed by a full media change with neuron maturation media. The neurons were maintained for 7-10 days before harvest for downstream targeted deep sequencing analysis and investigation of cell type on/off-target editing effects.

3.7 Targeted deep sequencing assays

3.7.1 On-target editing analysis (AmpSeq)

On-target Cas9-editing efficiency and checking for editing "leakage" in non-induced control samples was assessed as quality control of the dox-inducible CRISPR-Cas system, before proceeding to the final off-target editing analysis with the rhAmpSeq assay. On-target editing was analyzed with the standard targeted deep sequencing assay Amplicon-seq (AmpSeq) after harvesting the differentiated- and dox-induced cells. Here, a short, 250-350bp, dsDNA fragment within the target loci; *PCSK9*, *VEGFA2* or *HEK4* were amplified and barcoded, using a unique set of target-specific and barcoded AmpSeq primers, one forward (fw) and one reverse (rv). This strategy combines amplicon-specific barcoded primers with linked adapters for enabling multiplexed sequencing of a pool of several samples simultaneously and for facilitating downstream bioinformatics analysis. The protocol for preparing libraries for targeted deep sequencing was developed in-house by the AZ NGS team at the Translational Genomics Department (AstraZeneca, Gothenburg, Sweden) and it streamlines the sequencing protocol, reduces technical variation, and facilitates the

downstream data analysis of the sequencing results.

3.7.1.1 AmpSeq library preparation

Briefly, cells were harvested with either QuickExtract™DNA Extraction Solution ("QE", cat. nr.QE0905T, QE09050, Lucigen®Epicenter®) or with RLT Plus Lysis buffer ("RLT", cat. nr.79216, Qiagen) followed by DNA/RNA extraction using the Qiagen AllPrep®Mini- or Micro Kit (Cat. nr. 80204; nr. 80284, Qiagen). The extracted genomic DNA was prepared for targeted deep sequencing on the Illumina®NextSeq platform. In brief, an initial PCR program was run to amplify the region of interest, i.e. the CRISPR-Cas9-target cleavage site (*PCSK9*, *VEGFA2* or *HEK4*). Primer-ligation and DNA-amplification were performed with Phusion Hot Start II High-Fidelity PCR Master Mix (Cat. nr. F565S, ThermoScientific™) for samples with the specific U6-PCSK9 guide and with KAPA HiFi Hotstart Ready Mix (Cat. nr. 7958927001, Roche) for samples with the promiscuous U6-VEGFA2/U6-HEK4 guides. For each, in-house optimized PCR programs, respectively, were run.

PCR-amplified target sequences were verified by DNA electrophoresis on a 12-capillary Fragment Analyzer (Advanced Analytical Technologies). A smear analysis was performed to calculate the dilution factors for pooling the samples to a final volume of 100 µL, with the same initial fragment concentrations, using Qiagen Elution Buffer (Cat. nr.19086, Qiagen). The pooled sample was cleaned up with Agencourt AmPure XP beads (Beckman Coulter) and a secondary DNA electrophoresis was performed as a quality control prior to NGS. This was to verify the presence of the desired target sequences and no primer-dimer contamination. The final pool was diluted to 0.067 ng μL^{-1} in Qiagen EB and submitted for sequencing by the AZ NGS team on the Illumina®NextSeq500 platform. For more details consult *Appendix A*, Section A.5, and *Appendix B*, Section B.2 for a complete list of the primers and barcodes used for targeted deep sequencing (AmpSeq).

The sequencing output was analyzed using the AZ-internal NGS-analysis pipeline. Briefly, the reads were merged and mapped to the amplicon reference sequence by Burroughs-Wheeler Alignment (BWA), paired flash alignment to be exact. Variants (excluding the single nucleotide variations) were called with a minimum allele frequency of 0.0001% and a minimum base quality of 25 and maximum allowed read mismatch of 100. A maximum of 200 variants per amplicon were analyzed, where variants not overlapping with the cutting window by 8 nucleotides were excluded. The allowed variants were smaller (0-100 bp) insertions and deletions (indels) and small nucleotide polymorphisms/ substitutions (SNPs). Minimum read depth allowed was 10. Finally, the editing efficiencies were measured as the percentage of modified reads in mapped reads.

3.7.1.2 Genotyping of ODIn neurons

As the ODIn-PCSK9 culture in the first batch of differentiated neurons presented a comparatively low cell count upon reaching the Neuronal maturation stage, despite having seeded both cultures at the same initial seeding density, it was decided to genotype both cultures. A mix-up of the two cell lines was suspected, as the promiscuous guides (U6-VEGFA/U6-HEK4) were hypothesized to potentially cause

leakage-induced cytotoxic response and cell death, which could explain the low cell count. To assess whether the two ODIn-hiPSC-derived neuron cell lines had been mixed up, somewhere during differentiation, or in fact, were mixed cultures, genotyping was performed. Three different assays were used for this purpose; PCR, TIDE and AmpSeq. See *Appendix A*, Section A.4.6 for further details. A thorough investigation of where during cell line generation- or targeted differentiation this accidental mix-up occurred, was subsequently carried out. Targeted deep sequencing (AmpSeq) of the on-target loci was subsequently performed in all three cell lines of ODIn-hiPSCs, as well as in the two surviving cell lines of neuronal progenitors (after the third passage of neuronal induction), and in the mature dopaminergic neurons. No further genotyping assay of the other cell types (cardiomyocytes and hepatocytes) was performed.

3.7.2 Investigation of cell type effects on off-target editing with rhAmpSeq

Like the *VIVO study* [2], a semi-biased, two-step approach to CRISPR-Cas9 off-target editing analysis was taken. A CHANGE-seq *in vitro* assay was initially performed by A. Madsen and members of the NGS team at AZ to predict the most likely off-target cleavage sites for each RGN (PCSK9-SpCas9, VEGFA2-SpCas9, and HEK4-SpCas9) in a reference human genome. Targeted deep sequencing analysis was then performed to characterize and validate the predicted off-target effects in ODIn-hiPSC-derived cell lineages. As the off-target analysis was performed in three different cell lines, each with multiple cell lineages and replicates of both edited and unedited controls, an assay with higher throughput compared to standard amplicon sequencing was applied. Instead, for the off-target analysis, the rhAmpSeq approach was used.

3.7.2.1 rhAmpSeq panel design

rhAmpSeq panels were designed based on the identified off-target cleavage sites with the highest read counts in the CHANGE-seq assay, performed internally at AZ in collaboration between A. Madsen and the NGS team at the Translational Genomics Department. Three guide-specific CHANGE-seq assays with three biological replicates per guide were repeated twice, for optimization. The output was a spreadsheet with information about all the identified cleavage sites for each guide-specific CRISPR-SpCas9 nuclease. This included the off-target site annotation ($OT'X'$) and average read count, genomic coordinate for each site (chromosome, start and end position, and DNA strand), site sequence, site substitution number (number of mismatches allowed per target site), site sequence with gaps allowed, aligned target sequence and target site annotation. In total, between 200 and 6000 sites (*PCSK9*: 273, *HEK4*: 5193, and *VEGFA2*: 5819) were identified (including the on-target sites) for each guide RNA. As the input limit is $\geq$ 500 sites per indexed sample (guide RNA), the list of target sites for the rhAmpSeq panel had to be narrowed down. For this objective, for each guide, the identified sites were filtered and sorted based on read count. Next, the sites were cross-validated (checked for double

matches) with external references.

For the PCSK9-panel, the sites were compared to the results from an *in silico* prediction with the open-source and web-based version of CasOFFinder[6]. The input query contained the *PCSK9*-primer sequence (see *Appendix B*, Table B.1) and choice of PAM type (NGG for SpCas9). In addition, up to six mismatches between the guide sequence and the reference genome (human GRCh38/hg38) were allowed, as well as one RNA bulge and one DNA bulge. The output was a list of predicted genomic cleavage sites for PCSK9-SpCas9, with bulge type, DNA sequence, chromosome position, and direction, number of mismatches with the *PCSK9* spacer and target site, as well as bulge size. Out of the 273 cleavage sites, only 20 sites were found in the CasOFFinder output data. So, in addition to including the sites verified by CasOFFinder, all sites that were identified in more than one biological replicate of the CHANGE-seq assay were included in the rhAmpSeq panel. Furthermore sites with less than 4 mismatches, that could not be cross-verified with CasOFFinder or more than one replicate in the CHANGE-seq assay were also included. In total, 100 sites + the on-target site *PCSK9* were included in the rhAmpSeq panel for the PCSK9-guide.

For the VEGFA2 and HEK4 promiscuous guides, the cloud-version of CasOFFinder could only output a fraction of the total identified nucleolytic sites. Thus, it was decided to instead cross-validate their CHANGE-seq results with the sites identified *in vitro* in the original GUIDE-seq[103] and CIRCLE-seq[105] publications. This comparison generated panels of 155 potential off-target sites + 1 on-target site for the VEGFA2 guide and 123 off-target sites + 1 on-target site for HEK4. BED files (one per panel) containing information about the chromosome location, start- and end position of each site, were sent to IDT for the synthesis of rhAmpSeq primers. Out of the requested 101 sites for the PCSK9-panel, primers for 92 sites (incl. on-target site) could be synthesized. For VEGFA, 152 + the on-target site *VEGFA2* could be synthesized, and for HEK4 the final panel contained 120 OT sites + the on-target site *HEK4*.

3.7.2.2 rhAmpSeq library preparation

For the actual off-target analysis, the rhAmpSeq library preparation kit with customized guide-RNA specific primer panels from Integrated DNA Technologies (IDT) was used [89]. The library preparation was performed according to the manufacturer's instructions, for more details see *Appendix A* Section A.5.2.

In brief, the rhAmpSeq library preparation was performed twice as almost half of the prepared gDNA samples were below the required amount for pooling and NGS in the first round. The first prep was performed according to the manufacturer's instructions. However, for the second prep, two alterations were made to the original protocol. First, the PCR cycles were increased in both PCR1 (15 cycles instead of 10) and PCR2 (20 cyc. instead of 18) to increase the final amount of PCR product. Second, due to an issue with the Fragment Analyzer during the final quantitation of indexed PCR products, a second quantitation with another fragment analyzer was performed the day after library preparation. The output of the second FA was used for the calculation of final dilution factors and pooling. Furthermore, as some samples from the second prep were still below the accepted quantity for pooling and

NGS, but above the accepted limit in the first prep, these were used instead. So for the final pools, a mix of samples from the first and second library prep was used.

In total 63 samples of ODIn cell types were analyzed for off-target editing, 2-4 replicates per cell lineage, and dox-/non-dox treatment. Since the hepatocyte differentiation had to be rerun with a second differentiation protocol, this cell lineage was not included in the rhAmpSeq assay. The cell lineages assayed were, for ODIn-PCSK9: hiPSCs, cardiomyocytes and neurons (4 replicates per condition). For ODIn-VEGFA, hiPSC, cardiomyocytes, and neurons. For ODIn-HEK4, hiPSCs and cardiomyocytes were included in the rhAmpSeq assay. See Table A.10 in *Appendix A*, Section A.5.2 for an overview of the assayed samples per cell line.

3.7.2.3 rhAmpSeq analysis

For the final analysis of sequencing output, primary quality control and filtering of the raw data was performed by the NGS team. Read alignment was performed with BWA, paired flash alignment. Variants (excluding the single nucleotide variations) were called with a minimum allele frequency of 0.0001%, a minimum base quality of 25, and a maximum allowed read mismatch of 100. A maximum of 200 variants per amplicon were analyzed, where variants not overlapping with the cutting window by 8 nucleotides were excluded. The allowed variants were smaller (0-100 bp) insertions and deletions (indels) and single nucleotide polymorphisms/ substitutions (SNPs). The minimum read depth allowed was 10. Finally, the total editing efficiencies (indel + SNP) were measured as the percentage of modified reads out of the total number of mapped reads in each sample. The resulting output was three guide-specific data sets with mapped off-target sites and site-specific editing efficiencies for each replicate sample, respectively.

The rhAmpSeq data sets were further filtered for false positive mutations by increasing the cut-off threshold for read coverage to 500 reads per site. The parameter of interest was the mutation frequency, more specifically a mutation frequency over 0.1% (indicating that a site was cleaved by Cas9 at a confidence level above the NGS sensitivity limit). Thus, a binary analysis of the reported mutation frequencies for all samples was performed, where 0 = not edited (mut frequency $\leq$ 0.1%) and 1 = edited (mut frequency $\geq$ 0.1%). However, since some leakage-induced editing was detected in the control samples, the mutation frequencies for the dox-treated samples were normalized against the average mutation frequency of the control replicates, for each condition. Any negative values were also assigned '0' as it implied leakage-mediated editing in potentially premature cell types. These edits fall outside the scope of investigating cell type-specific effects on off-target editing. Rather, only the mature and dox-treated samples were assessed for differential off-target editing. The normalized mutation frequencies for the dox-treated specimens were then sorted with binary analysis and visualized with heatmaps.

4

Results

4.1 hiPSC differentiation and cell characterization assays

Human iPSCs were used as a starting point, where inducible SpCas9 had been integrated into the *AAVS1* locus [66]. Additionally, the Xential-based method[64] had been employed to introduce a gRNA cassette into the *HBEGF* locus. A. Madsen demonstrated that this process resulted in the successful generation of cell lines for a specific guide, U6-PCSK9 (PCSK9), as well as two promiscuous guides, U6-VEGFA2 (VEGFA2) and U6-HEK4 (HEK4).

4.1.1 hiPSC evaluation and trilineage differentiation

4.1.1.1 qPCR

ODIn-hiPSC pluripotency was initially validated with qPCR targeting pluripotency marker genes *POU5F1* and *NANOG* on mRNA from undifferentiated ODIn-hiPSCs (Figure 4.1). The readout from qPCR, marker-specific C_T values, were analyzed with the C_T-method, except for normalization against control samples, due to the absence of suitable reference samples. The undifferentiated ODIn-hiPSCs showed no significant expression of the *NANOG* marker after normalization against both housekeeping genes, *PGK1* and *GAPDH*, and was therefore excluded from the analysis. The hiPSCs did however show positive fold-change expression (1-87x) and thus upregulation of the *POU5F1* gene in all three cell lines, indicating variation, but definitive evidence of pluripotency across ODIn-hiPSC lines (Figure 4.1). Note however that only one biological replicate (x3 technical replicates) per cell line was assayed, which reduces the statistical significance of reported fold-changes for in the hiPSC qPCR assay.

Thereafter, trilineage differentiation potential was evaluated by qPCR and subsequent C_T-analysis on mRNA from trilineage differentiated ODIn-hiPSCs. For this purpose 2-3 germ layer-specific marker genes per lineage were assayed; endoderm (*SOX17 CXCR4* & FOXA2), mesoderm (*Brachyury (T)*, *NCAM* & *CXCR4*) and ectoderm (*NCAM* & *Nestin*). The normalized readout $2^{-\Delta\Delta C_T}$ values indicated overall upregulation of the trilineage-specific marker genes in the range of 1-20000 fold-change, respectively (Figure 4.2). The least expressed markers were *NCAM* (mesoderm) and *Nestin* (ectoderm), in all three cell lines, with fold-changes ranging between 0.5-10x (ODIn-PCSK9 and ODIn-HEK4) and 90-130x (ODIn-VEGFA2).

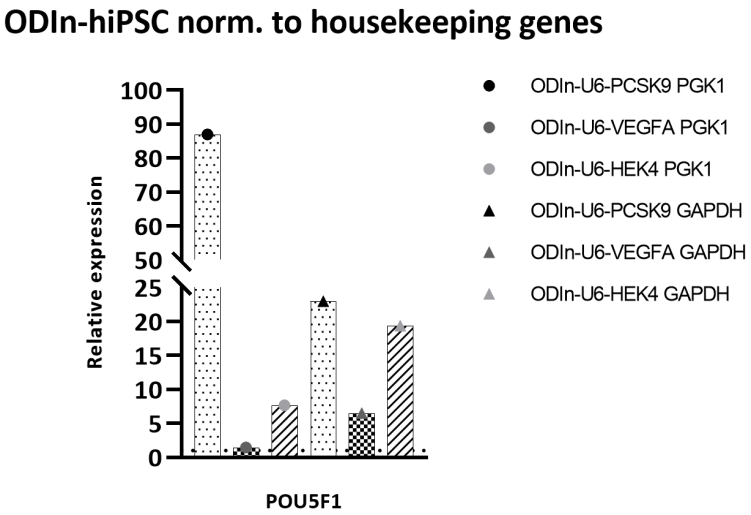
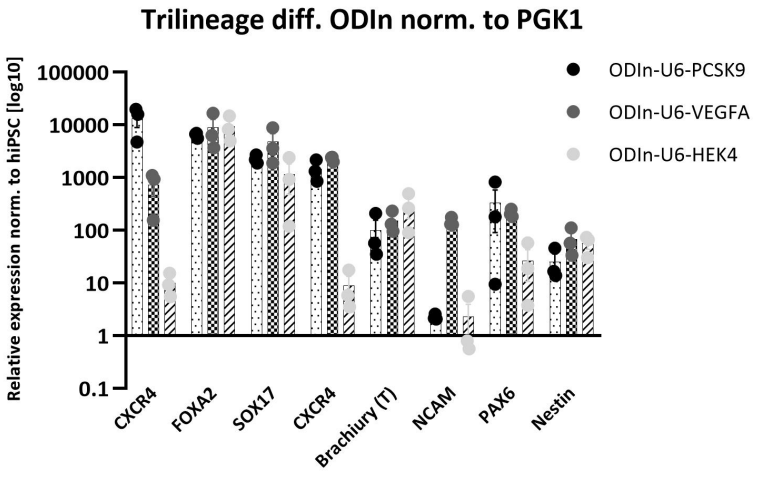


Figure 4.1: Relative expression of the *POU5F1* pluripotency marker gene in all three ODIn-hiPSC lines. Measured C_T-values were normalized against the two housekeeping genes, *PGK1* (dot), and *GAPDH* (triangle), respectively, and presented as comparative fold-changes.

Except for these two markers, comparatively similar expression levels of CM- markers and the pluripotency marker were observed across all cell lines and batches. Nevertheless, some variability was recognized between replicates in all three cell lines. Still, the qPCR assays indicated that the ODIn-hiPSCs possess the ability to differentiate into cell types in all three germ layers.

(a)



(b)

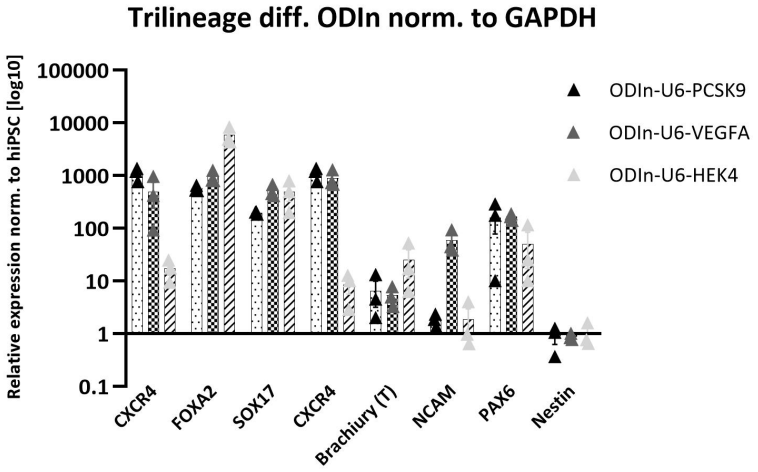


Figure 4.2: Relative expression (fold-change) of lineage-specific marker genes in endodermal and mesodermal cells at Day 5 of differentiation, and in ectodermal cells at Day 7 of differentiation. Measured C_T-values were normalized against the equivalent expression in hiPSCs (each data point), and against the expression of housekeeping genes, (a) *PGK1*, and (b) *GAPDH*, respectively. Mean $2^{-\Delta\Delta C_T}$ values are presented for each biological replicate on a log10-scale, with variability expressed as SEM and depicted with error bars.

4.1.1.2 Immunostaining and flow cytometry analysis

The pluripotency characteristics of the ODIn-hIPSCs were further evaluated by staining the cells maintained in the DEF-CS System (Takara Bio), with the pluripotency-associated cell surface marker *stage-specific embryonic antigen 4* SSEA4 (Miltenyi Biotech). To check for unspecific binding, an isotype control (IgG) stain was also performed, and for comparison, a positive control (ODIn-hiPSCs without sgRNA cassette) stained for both SSEA4 and IgG was included as well. Flow cytometry analysis of SSEA4 expression showed $\geq$ 95% of hiPSCs positive for SSEA4 in all

three cell lines, indicating retained pluripotency after cell line generation, see Figure 4.3. Together, these two assays provided a strong indication for successful trilineage differentiation and validated the ODIn-hiPCs as highly pluripotent. For additional figures with control conditions, see Figure D.1 in *Appendix D*, Section D.1.

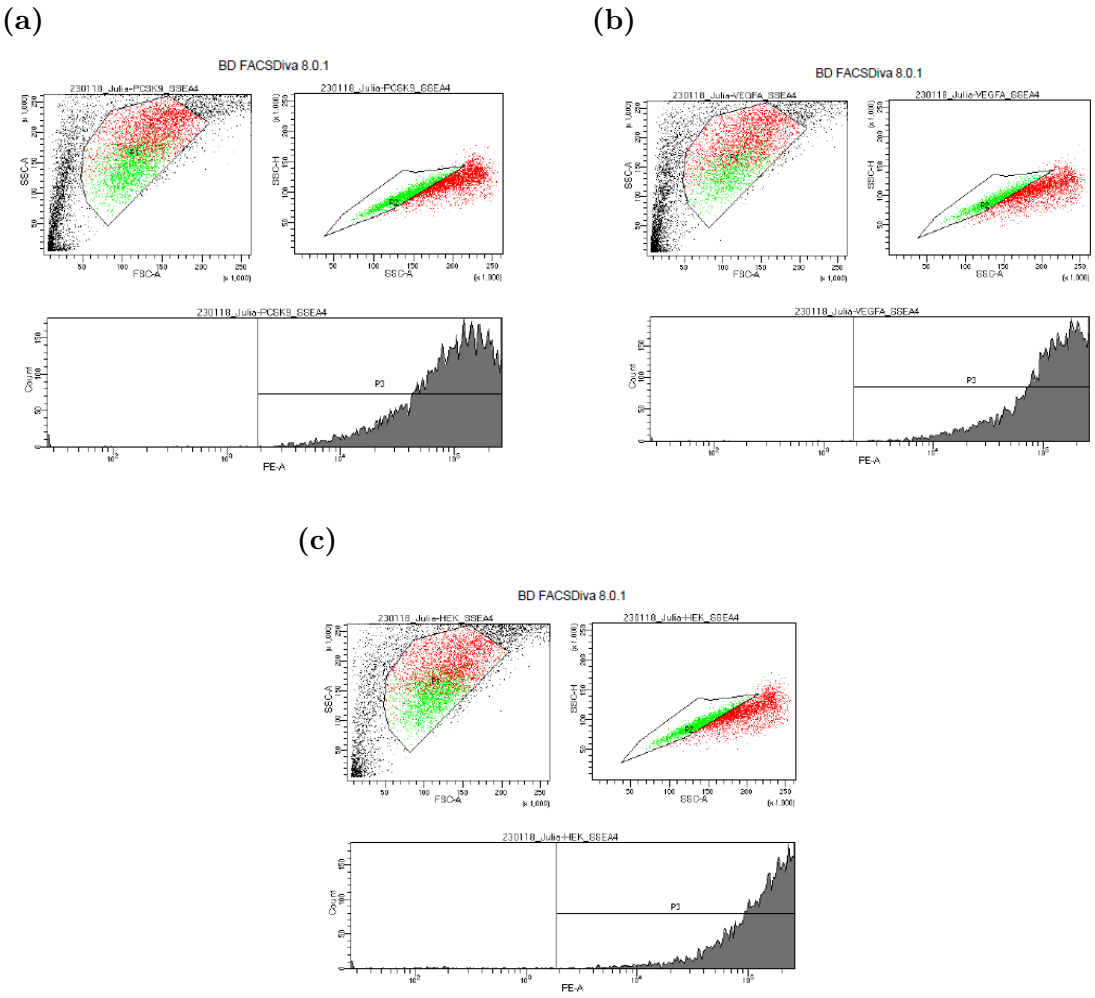


Figure 4.3: Flow cytometry analysis of the three ODI-hiPSC lines for pluripotency validation, isotype control stain (IgG-PE) and pluripotency marker SSEA4 (SSEA4-PE). (a) SSEA4 stained ODIn-PCSK9 hiPSCs, (b) SSEA4 stained ODIn-VEGFA2 hiPSCs, and (c) SSEA4 stained ODIn-HEK4 hiPSCs.

4.1.2 Evaluation of cell type-directed differentiation

During hepatic differentiation with the Takara Bio-protocol and neuronal differentiation with the STEMCELL Tech. protocol, one of the cell lines with promiscuous guides died, each. For the hepatic diff, it was the ODIn-U6-VEGFA line (at day 4-5 of definitive endoderm diff.), and for the neurons, it was the ODIn-HEK4 cell line (at neuronal induction stage). Possible explanations for this are reviewed in the *Discussion*.

4.1.2.1 Cell morphology

The ODIn-hiPSCs possessed typical iPSC morphology, as described by the Cellartis®DEF-CS™System User Manual (Takara bio). The hiPSCs grew in a relatively homogeneous monolayer with undifferentiated cells and presented an iPSC-typical high nucleus-to-cytoplasm ratio and defined cell borders (Figure 4.1, subfigure (i)). Contrarily, the ODIn-hiPSC-derived cell lineages presented a very different cell morphology, in line with visual descriptions presented in each of the respective differentiation kits.

The cardiomyocytes presented a morphology of alternating cell-free regions and regions of dense clusters with somewhat overlapping cells covered by a dense (collagen-rich) extracellular matrix (Figure 4.1, subfigures (a)-(c)). Most prominently, the dense cell-covered regions were beating more or less synchronously from Day 10 onwards, with some regions beating already at Day 8 of the differentiation protocol. The ODIn-PCSK9 and ODIn-VEGFA2 cardiomyocytes presented the most cardiomyocyte-like features, in terms of the number of synchronously beating cell clusters across replicate samples. The ODIn-HEK4 CMs presented similar topological dispersion of cells into clusters, but very few to none of these clusters showed pronounced beating, across both batches.

The ODIn-hepatocytes derived with the Cellartis®kit (Figure 4.1, subfigures (d) & (e)), were covered by a gelatinous overlay and thus more difficult to assess morphology-wise. Still, hepatocyte-like features with hexagonally-shaped and densely clustered cells could still be recognized through the overlay, of both ODIn-HEK4 (d) and ODIn-PCSK9 (e) hepatocytes (batch I). Nevertheless, the STEMCELL tech. kit generated hepatocytes with superior hepatocyte-like features (Figure 4.1, subfigure (f)). Note that in this figure, only the ODIn-PCSK9 hepatocytes are represented, but similar morphology was observed in the ODIn-VEGA2 and ODIn-HEK4 cultures, respectively. The second batch of hepatocytes also presented more consistent features across replicates, compared to the first batch.

Both surviving cell lines of ODIn-neurons showed neuron-like features of bipolar cells with long branching axons and smaller branching dendrites (Figure 4.1, subfigures (g) and (h)). Note that in the figure, the cell morphology of the first batch of ODIn-VEGFA2 neurons (g) and the second batch of ODIn-PCSK9 neurons (h) are presented, respectively. In both batches, a confounding effect to downstream characterization was that the cell count of the ODIn-PCSK9 neurons was considerably lower than that of the ODIn-VEGFA2 neurons, despite being consistently seeded at the same initial seeding density. Note furthermore that the ODIn-VEGFA2 neurons were later identified to be a mixed culture of both cell lines, see Section 4.2.1.1 for further explanation of this and downstream implications.

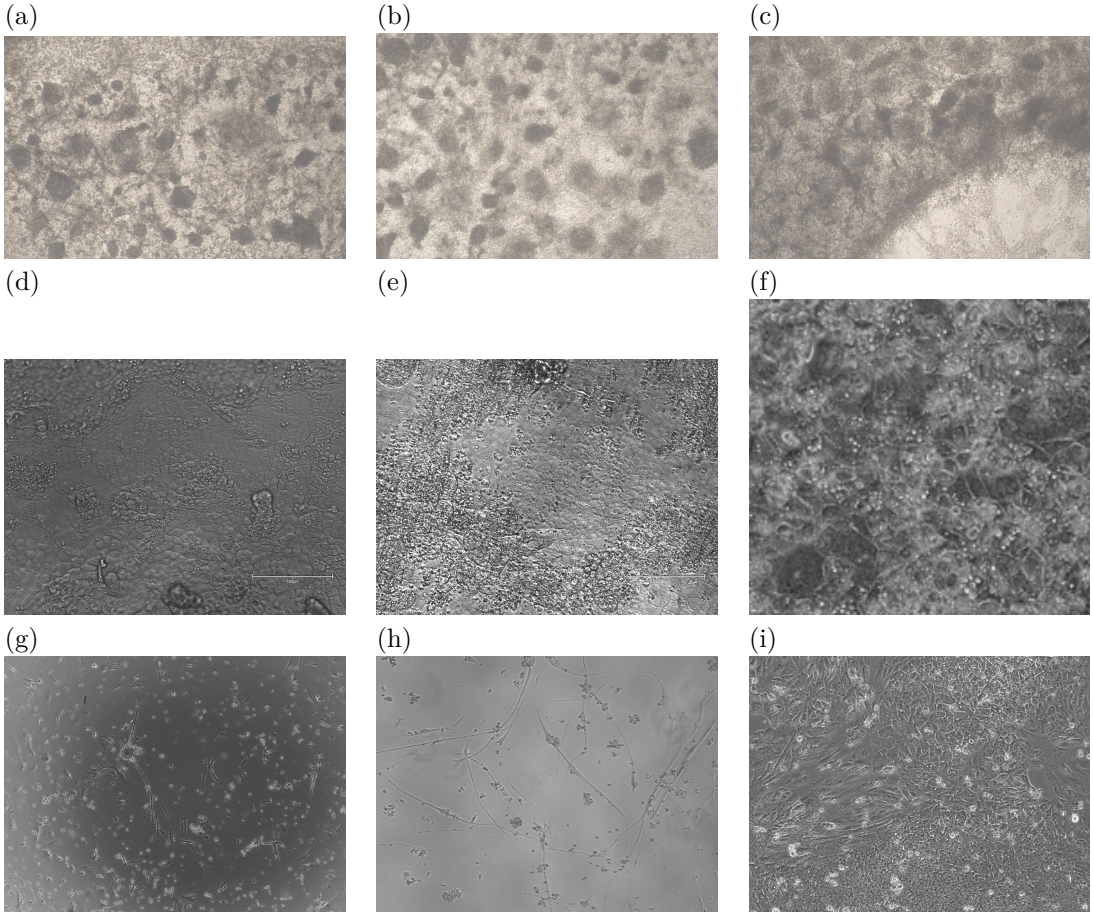


Table 4.1: Microscopy images of ODIn-hiPSCs and downstream cell lineages at progenitor and mature cell stages, all images are taken at 10x magnification. (a) ODIn-PCSK9 mature ventricular cardiomyocytes, (b) ODIn-VEGFA2 mature ventricular cardiomyocytes, (c) ODIn-HEK4 mature ventricular cardiomyocytes (all in STEMCELL™Tech. CM kit), (d) ODIn-HEK4 mature hepatocytes, (e) ODIn-PCSK9 mature hepatocytes (both Takara Bio hep kit), (f) ODIn-PCSK9 mature hepatocytes (STEMCELL Tech. hep kit), (g) ODIn-VEGFA2 mature neurons, (h) ODIn-PCSK9 mature neurons (both STEMCELL™Tech. nr kit), and finally (i) ODIn-hiPSCs (without sgRNA) in DEF-CS™System (Takara Bio).

4.1.2.2 qPCR

To characterize the differentiated cell types and verify successful differentiation, comparative gene expression analysis by quantitative PCR (qPCR) was performed. Table 4.2 presents an overview of the expected versus actual fold-change of cell type-specific gene expression, as demonstrated by qPCR analysis. Note that the actual positive fold-change of the *TUBB3* applies only to the ODIn-VEGFA2 neurons. The ODIn-PCSK9 neurons demonstrated a negative fold-change expression of *TUBB3* after normalization against the equivalent hiPSC expression and against the expression of the housekeeping genes *PGK1* and *GAPDH*, respectively.

Table 4.2: Expected and experimental regulation of cell-type specific genetic markers (mRNA) in differentiated cell types (all cell lines) when normalized against equivalent mRNA levels in ODIn-hiPSCs and against the expression of the housekeeping genes *GAPDH* and *PGK1*, respectively. + = upregulation, - = downregulation, and N/A = 'not assigned'.

Sample	**Genetic marker targeted**	**Expected fold-change (mean $2^{-\Delta\Delta C_T}$)**	**Actual fold-change (mean $2^{-\Delta\Delta C_T}$)**
Cardiomyocytes (ventricular)	*MYH6*, *NKX2-5*, *TNNT2*, and *POU5F1*	+++-	+++
Hepatocytes (Cellartis®)	*HNFA4*, *A1AT*, *ALB*, and *POU5F1*	+++-	++N/A-
Hepatocytes (STEMdiff™)	*HNF4A*, *A1AT*, *ALB*, and *POU5F1*	+++-	N/A+N/A-
Neurons (midbrain dopaminergic)	*TUBB3*, *FOXA2*, *TH*, and *POU5F1*	+++-	+++-

Reviewing the results from the C_T-analysis for quantifying the relative expression of each cell type and cell line, in the order they were assayed, the first cell type to be analyzed was the ventricular cardiomyocytes. Two batches of ODIn-hiPSC-derived CMs were generated with the same diff. kit (STEMCELL Tech.), but with minor adjustments to the second assay for optimization purposes. Figure 4.4 illustrates the normalized fold-changes ($2^{-\Delta\Delta C_T}$ values) for each genetic marker in the CM samples. Overall, both batches of CMs showed positive fold-changes and thus upregulation of the assayed cardiomyocyte-specific markers (*MYH6*, *NKX2-5* and *TNNT2*). Moreover, all cell lines showed negative fold-changes and thereby downregulation of the pluripotency marker *POU5F1* (Figure 4.4). The most prominent fold-change expression was observed in both batches of ODIn-PCSK9 CMs (dotted bars) and ODIn-VEGFA2 CMs (chess-patterned bars), where fold-changes were *geq* 100x for the CM-markers and *leq* 0.01 for *POU5F1*. These results are in line with the pronounced cardiomyocyte-like features observed with microscopy, that is, widespread synchronous beating, across both batches of ODIn-PCSK9 and ODIn-VEGFA2 CMs. Notably, the normalized expression of *TNNT2* was not as high in either of the cell lines compared to the other CM-markers (1-10x fold-change). For the ODIn-PCSK9 and ODIn-VEGFA2 samples, the expression of all CM-markers appeared to be slightly higher in the second batch of cardiomyocytes, whereas the opposite was observed for the ODIn-HEK4 cell line (striped bars).

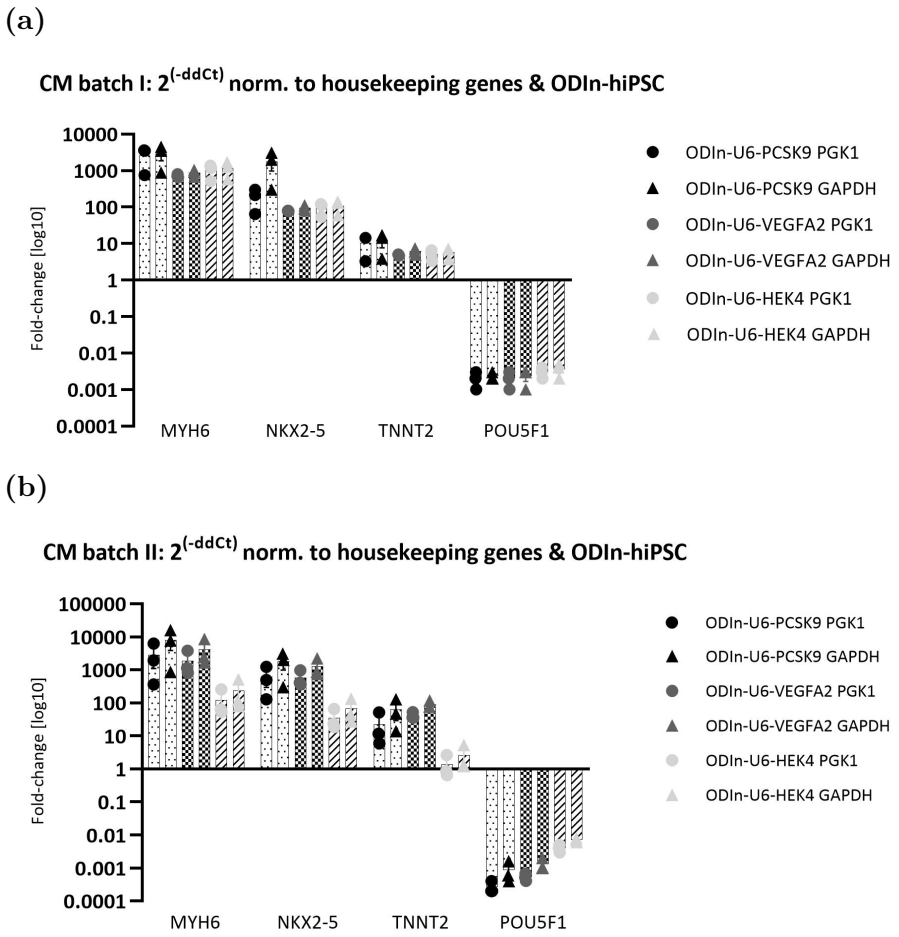


Figure 4.4: Relative expression (fold-change) of cardiomyocyte-specific marker genes (*MYH6*, *NKX2-5* and *TNNT2*) and *POU5F1*-pluripotency marker gene in ODIn-hiPSC derived cardiomyocytes, (a) batch I CMs, (b) batch II CMs. Measured C_T-values were normalized against the equivalent expression in hiPSCs (each data point), and against the expression of housekeeping genes *PGK1* (dot) and *GAPDH* (triangle), respectively. Mean $2^{-\Delta\Delta C_T}$ values are presented for each biological replicate on a log10-scale, with variability expressed as SEM and depicted with error bars.

For qPCR validation of hepatic differentiation with both protocols (batch I: Takara kit & batch II: STEMCELL tech. kit), the expression of hepatocyte-specific genetic markers *HNF4A*, *A1AT* and *ALB*, as well as the pluripotency-marker *POU5F1*, were analyzed with qPCR. Figure 4.5 visualizes the C_T-analysis results for both ODIn-hepatocyte batches. *ALB*-expression could not be quantitatively determined (no readout) in the hiPSC controls across all cell lines, and batches. Furthermore, *HNF4A* gave no readout in the hiPSC controls in the qPCR assay of the second batch of hepatocytes. The undermined values, likely due to too low gene expression, prevented the use of the $2^{-\Delta\Delta C_T}$ method for comparative gene expression analysis of the *ALB* markers, and *HNF4A* markers, respectively, in affected hepatocyte samples. The hepatic markers that could be quantified in both hiPSCs and hep-

atocytes, *HNF4A* (batch I - Figure 4.5, subfigure (a)) and *A1AT* (both batches), had positive fold-changes in all assayed ODIn hepatocytes and can therefore confidently be assessed as upregulated. However, in batch I, the results was extremely variable (several 10x), both between replicates and between the two cell lines. The ODIn-HEK4 specimens in the first hepatocyte batch (Figure 4.5, subfigure (a), striped bars) had considerably higher expression of *HNF4A* and *A1AT* compared to the ODIn-PCSK9 specimens (10x, indicating overall variability in differentiation outcomes between cell lines. The second batch of ODIn-hepatocytes (Figure 4.5, subfigure (b)) presented less replicate- and cell line variability in fold-changes (*leq* 10x), and overall larger fold-change expression for each marker (several 10-folds).

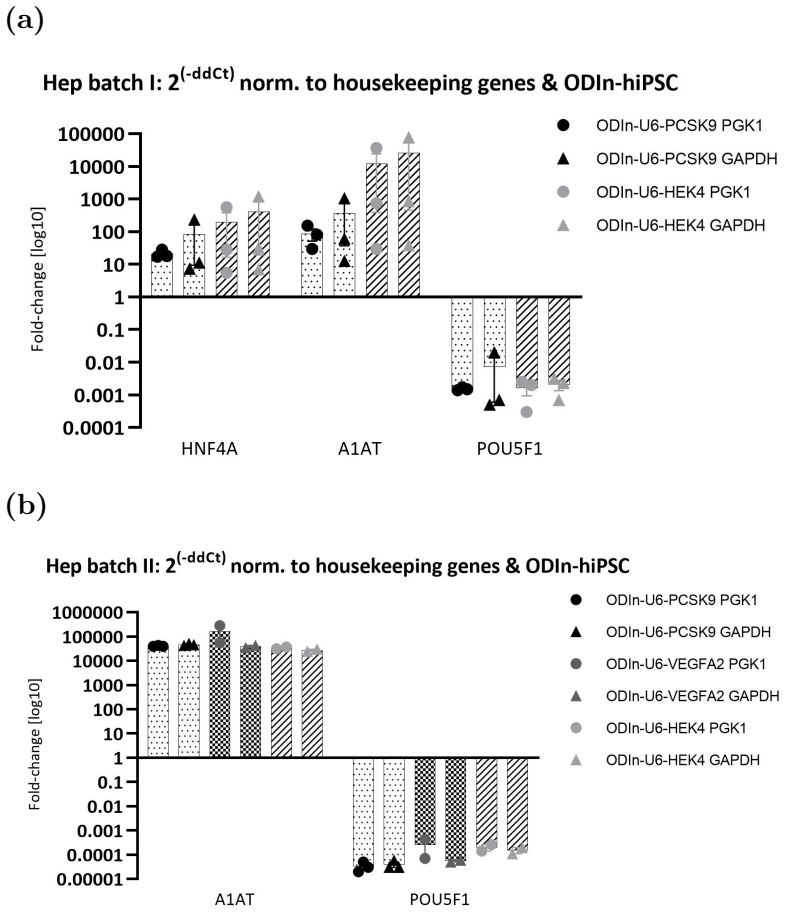


Figure 4.5: Relative gene expression of hepatocyte-specific markers (*HNF4A* and *A1AT*), and *POU5F1*-pluripotency marker in ODIn-hiPSC derived hepatocytes; (a) batch I hepatocytes (Takara Bio kit), and (b) batch II hepatocytes (STEMCELL tech. kit). C_T-values were normalized against the equivalent expression in hiPSCs (each data point), and against the expression of housekeeping genes *PGK1* (dot) and *GAPDH* (triangle), respectively. Mean $2^{-\Delta\Delta C_T}$ values are presented for each biological replicate on a log10-scale, with variability expressed as SEM and depicted with error bars.

For the qPCR characterization of the ODIn-hiPSC-derived neurons, comparative expression of the three different midbrain-dopaminergic marker genes (*TUBB3*, *TH*,

and *FOXA2*), as well as *POU5F1*, were analyzed with qPCR in the two surviving neuronal cell lines (ODIn-PCSK9 and ODIn-VEGFA2). The first batch of neurons generated inconclusive qPCR results, that is, all genetic markers except for the two housekeeping genes *PGK1* and *GAPDH* gave no quantitative gene expression readout (C_T-values). Most likely, this was due to a low output of purified and intact RNA (approximately 3 ng μL^{-1}), as indicated by Nanodrop measurements. Batch II generated a larger amount of mRNA (approximately 12 ng μL^{-1}), which could be quantified with PCR. The results of $2^{-\Delta\Delta C_T}$ comparative gene expression analysis in the second batch of neurons, is illustrated in Figure 4.6. Notably, the *FOXA2*-marker gave no quantitative readout (no C_T value) in the ODIn-hiPSC samples. Similar to the hepatocyte qPCR-assay, this prevented the use of the $2^{-\Delta\Delta C_T}$-method for comparative gene expression analysis of this marker. The three remaining markers were successfully quantified in both neurons and hiPSCs. Overall, the ODIn-PCSK9 culture (Figure 4.6, dotted bars) presented a lower expression of the neuron-specific markers, compared to the ODIn-VEGFA2 culture (Figure 4.6, chess-patterned bars). $2^{-\Delta\Delta C_T}$ analysis showed positive fold-change gene expression after normalization of *TUBB3*; 1-10x fold-change in ODIn-VEGFA2 nr, and *TH*; 10-100x, and 100-1000x fold-change expression in ODIn-PCSK9 nr, and ODIn-VEGFA2 nr, respectively. In contrast, the ODIn-PCSK9 neurons showed a negative fold-change of *TUBB3* after normalization (0.05x). This indicates less significant *TUBB3* expression compared to the expression of the two housekeeping genes. Nevertheless, *POU5F1* was significantly downregulated in both neuronal cell lines, indicating an overall loss of pluripotency.

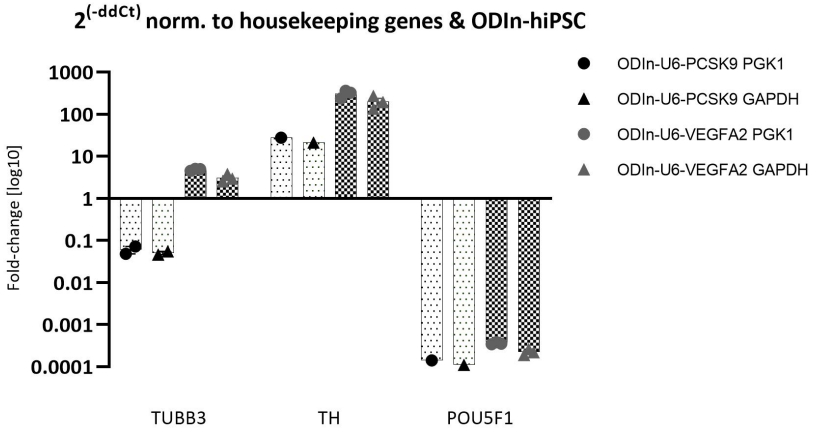


Figure 4.6: Relative gene expression of neuron-specific marker genes (TUBB3, TH, and *FOXA2*) and *POU5F1*-pluripotency markers for the second batch of ODIn-hiPSC derived neurons; ODIn-PCSK9 (dotted bars) and ODIn-VEGFA2 (chess-patterned bars). C_T values were normalized against the equivalent expression in hiPSCs (each data point), and against the expression of housekeeping genes *PGK1* (dot) and *GAPDH* (triangle), respectively. Mean $2^{-\Delta\Delta C_T}$ values are presented for each biological replicate on a log10-scale, with variability expressed as SEM and depicted with error bars.

4.1.2.3 Immunostaining and flow cytometry analysis

The proportion of mature ventricular cardiomyocytes out of the total population of cardiomyocyte-differentiated cells was assessed using an anti-cardiac-troponin T (cTnT) antibody and analyzed with flow cytometry. Flow cytometric acquisition of the first batch of ODIn-hiPSC-derived cardiomyocytes revealed at most 34% cTnT positive cells (ODIn-U6-VEGFA). ODIn-HEK4 showed 13% cTnT-positive cells and ODIn-PCSK9 presented the least amount of cTnT-positive cells (9%), see Figure 4.7. Note further that the first batch of cardiomyocytes was not stained for any pluripotency marker nor include any positive control cell line. For additional figures with control conditions, see Figure D.2 in *Appendix D*, Section D.1.

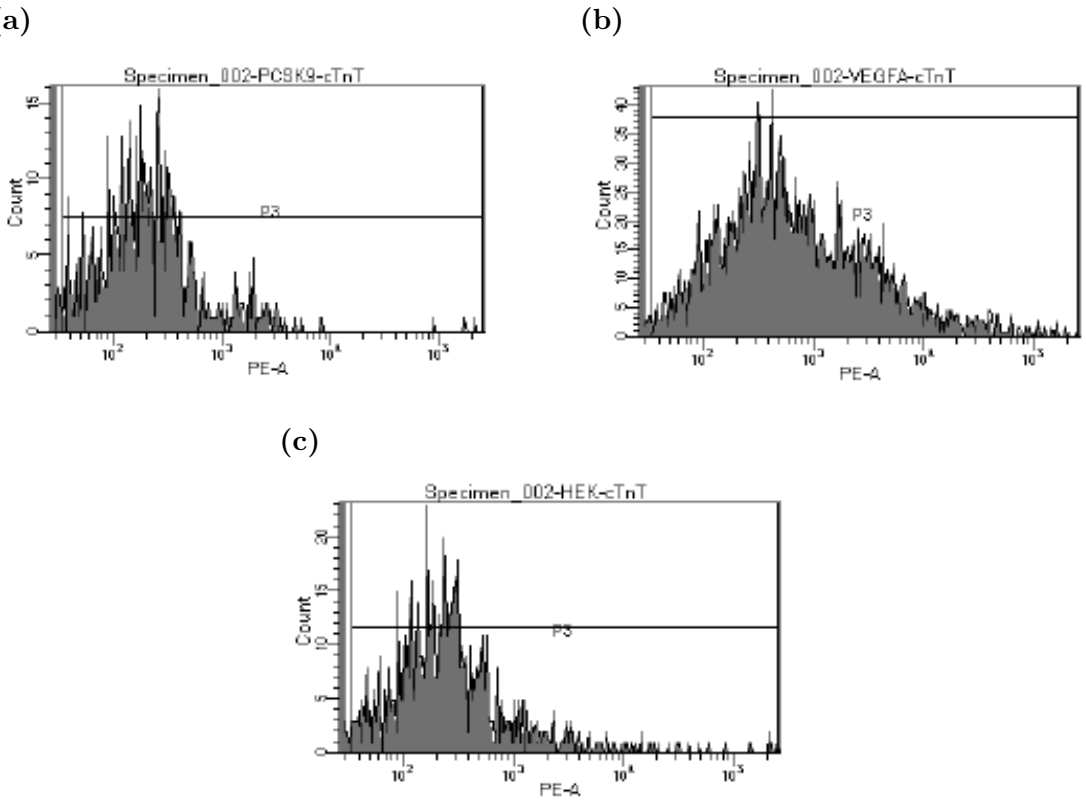


Figure 4.7: Flow cytometry analysis results of the first batch of ODIn-hiPSC-derived cardiomyocytes generated with the STEMCELL Tech. kit, (a) cTnT stained ODIn-PCSK9, (b) cTnT stained ODIn-VEGFA2, and (c) cTnT stained ODIn-HEK4. All cell lines were stained with PE-conjugated Cardiac Troponin T (cTnT-PE) and an isotype control stain (IgG-PE), except for ODIn-PCSK9 (no IgG stain) due to low cell count at harvest. One biological replicate per stain was assayed and the subpopulation of cTnT-positive cells was gated based on forward- and side scatter (FSC/SSC), removing any cell duplets and smaller particles.

For the flow cytometry analysis of the second cardiomyocyte batch (optimized STEMCELL Tech. kit protocol), a PE-conjugated SSEA4 antibody was also included, besides cTnT-PE and IgG-PE. The purpose of the SSEA4-stain was to assess the proportion of (undesired) pluripotent stem cells/progenitor cells remaining in

the CM culture at harvest. Moreover, the osteosarcoma-derived cell line U2-OS was tested as a positive control for the expression of cTnT, but presented no significant expression of this marker and was therefore excluded from the figure (Figure 4.8). As depicted in subfigure (b), ODIn-VEGFA2 CMs reported the lowest expression of cTnT (17%) when looking exclusively at singlets (gate P4 in the figure). The corresponding percentage of cTnT-positive CMs for ODIn-PCSK9 (a) and ODIn-HEK4 (c) were 23% and 25%, respectively. When including doublets, the proportion of cTnT-positive cells increased to 31%, 35%, and 31%. Meanwhile, 20% of ODIn-PCSK9 (d), 16% of ODIn-VEGFA2 (e), and 27% of ODIn-HEK4 (f) CMs stained positive for SSEA4, respectively. For additional figures with IgG-control stains, see Figure D.3 in *Appendix D*, Section D.1.

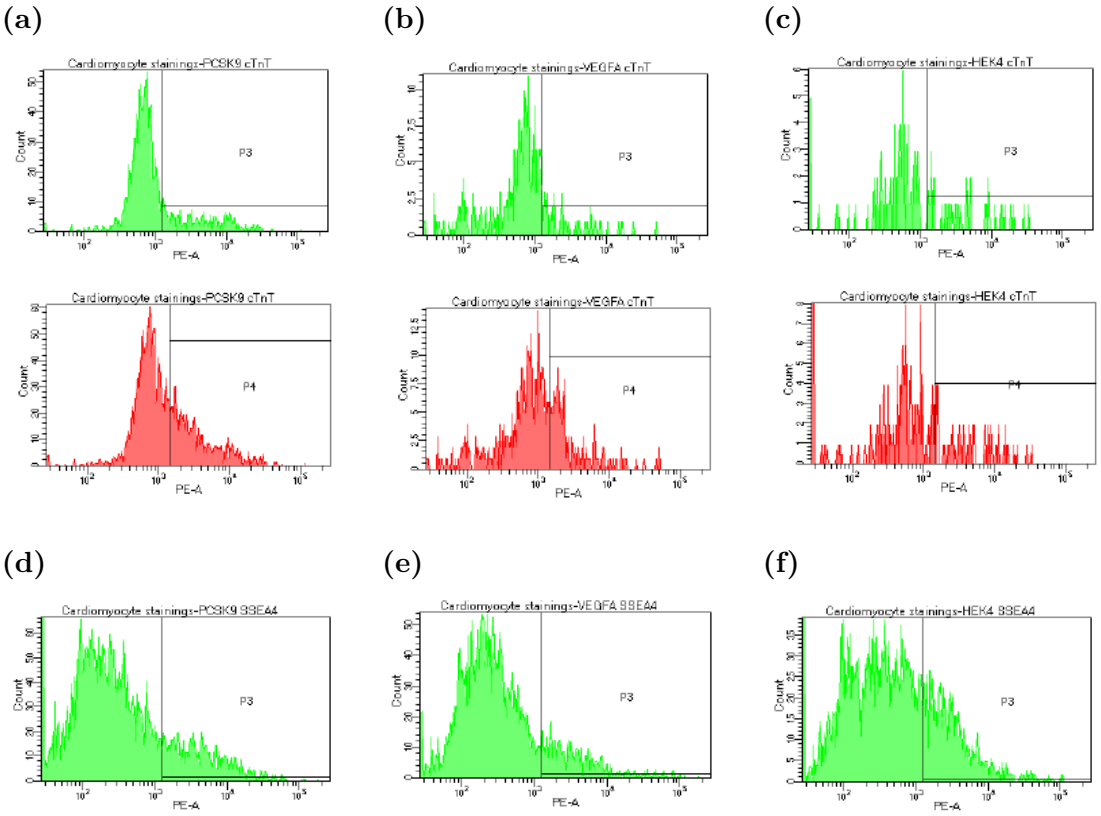


Figure 4.8: Flow cytometry analysis results of the second batch of ODIn-hiPSC-derived cardiomyocytes: (a) cTnT stained ODIn-PCSK9, (b) cTnT stained ODIn-VEGFA2, (c) cTnT stained ODIn-HEK4, (d) SSEA4 stained ODIn-PCSK9, (e) SSEA4 stained ODIn-VEGFA2, and (f) SSEA4 stained ODIn-HEK4. One biological replicate per stain was assayed, and the subpopulation of cTnT-positive cells was gated based on forward- and side scatter (FSC/SSC), removing any cell duplets and smaller particles.

Next, the two hepatocyte cultures (ODIn-PCSK9 and ODIn-HEK4) derived with the Takara Bio-protocol (batch I) were assessed with flow cytometry. Both cell lines were stained with a PE-conjugated antibody targeting the asialoglycoprotein receptor 1 (ASGPR1-PE), see results in Figure 4.9, subfigures (a) ODIn-PCSK9

and (c) ODIn-HEK4. Only ODIn-PCSK9 was further stained for the pluripotency marker SSEA4 (subfigure (b)), due to too low cell count at harvest for the ODIn-HEK4 culture. Also, the hepatocellular carcinoma cell line HepG2 was assayed as a positive control for the expression of ASGPR1 and stained close to 100% ASGPR1-positive (96%). In contrast, only 5% of ODIn-PCSK9 hepatocytes and 14% of ODIn-HEK4 cells stained positive for ASGPR1 (subfigures (a), and (c), respectively). Meanwhile, 10% of ODIn-PCSK9 were positive for SSEA4 (subfigure (b)). For additional figures with HEPG2- and IgG-control stains, see Figure D.4 in *Appendix D*, Section D.1.

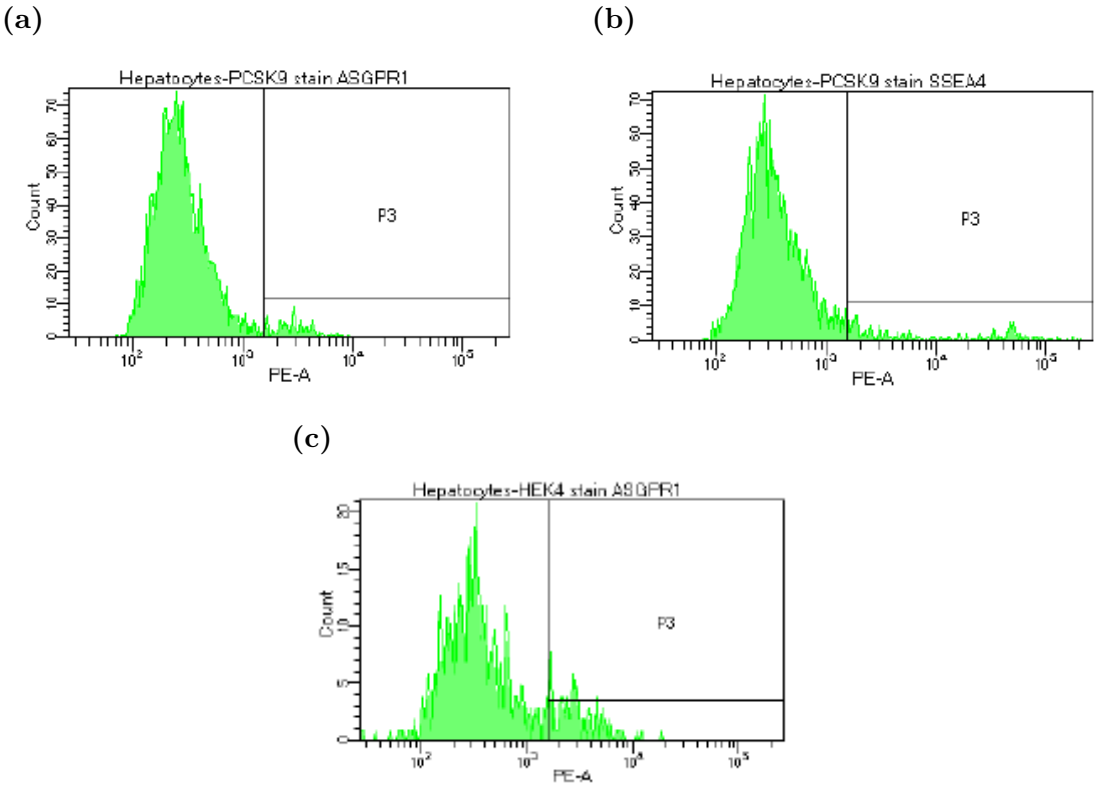


Figure 4.9: Flow cytometry analysis results of the two surviving cell lines of ODIn-hiPSC-derived hepatocytes with the Cellartis®Takara Bio kit, (a) ASGPR1 stained ODIn-PCSK9, (b) SSEA4 stained ODIn-PCSK9, and (c) ASGPR1 stained ODIn-HEK4. Both cell lines were stained with a PE-conjugated antibody targeting asialoglycoprotein receptor 1 (ASGPR1-PE), but only ODIn-PCSK9 was stained for the isotype control (IgG-PE) and the pluripotency marker SSEA4 (SSEA4-PE). The hepatocellular carcinoma cell line HepG2 was assayed as a positive control for the expression of albumin, as well as for the isotype control stain (IgG). One biological replicate per stain was assayed and the subpopulation of ASGPR1-positive cells was gated based on forward- and side scatter (FSC/SSC), removing any cell duplets and smaller particles.

For the second batch of hepatocytes, the three ODIn hepatocyte cell lines were stained with albumin (ALB-FITC) instead of ASGPR1, in line with recommendations from the manufacturer of the kit (STEMCELL Tech.). The hepatocellular carcinoma cell line HepG2 was also successfully assayed as a positive control for the expression of albumin but tested only 20% positive for this marker, in contrast to the ASGPR1-stain where the same HepG2 control stained 96% positive. As presented in Figure 4.10, subfigures (a)-(c), the ODIn hepatocytes also presented a low proportion of albumin-positive cells, with only 2-8% albumin-positive across all three cell lines. Note that none of the ODIn hepatocytes in the second batch were stained for the pluripotency marker SSEA4, due to low cell count at harvest. For additional figures with HePG2- and IgG-control conditions, see Figure D.5 in *Appendix D*, Section D.1.

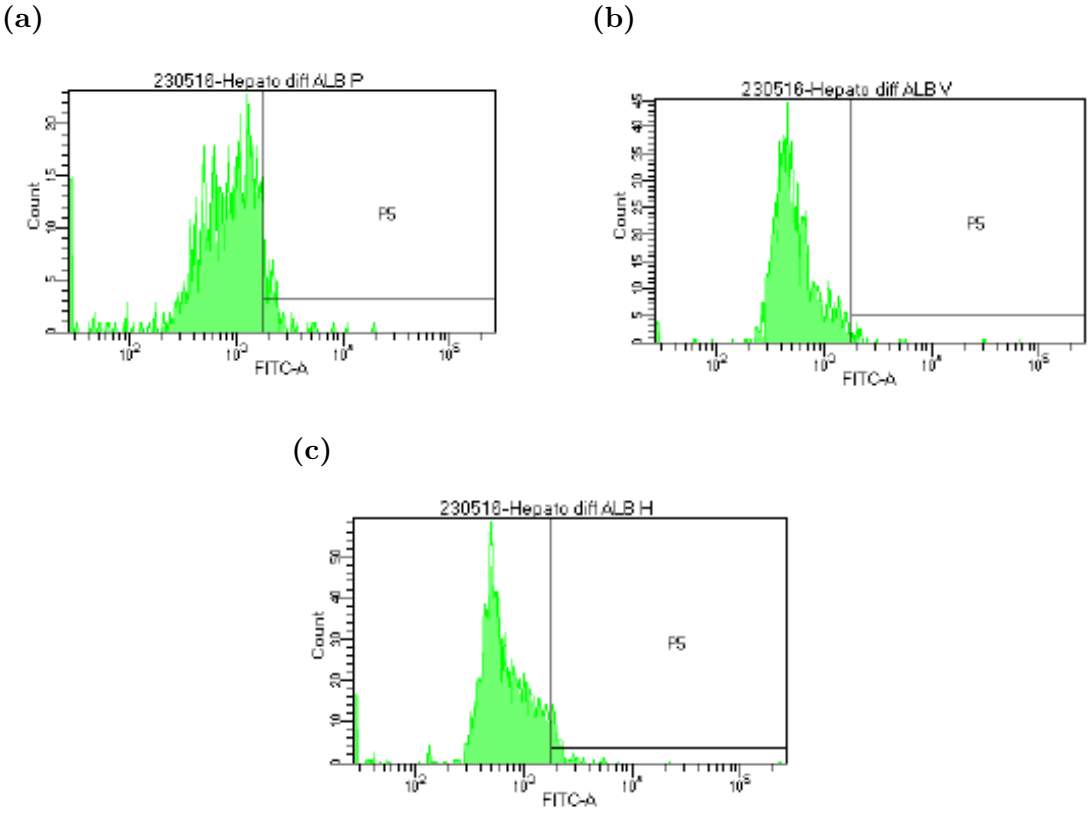


Figure 4.10: Flow cytometry analysis results of the second batch of hepatocytes generated with the STEMCELL Tech. kit, (a) ALB-stained ODIn-PCSK9, (b) ALB-stained ODIn-VEGFA2, and (c) ALB-stained ODIn-HEK4. These hepatocytes were stained with FITC-conjugated albumin (ALB-PE) instead of asialoglycoprotein receptor 1 (ASGPR1-PE), as well as an isotype control stain (IgG-FITC) for the HepG2 control cell line (exclusively). One biological replicate per stain was assayed and the subpopulation of ALB-positive cells was gated based on forward- and side scatter (FSC/SSC), removing any cell duplets and smaller particles.

Finally, as illustrated by Figure 4.11, subfigure (b), close to 90% of ODIn-VEGFA2 neurons from the first batch stained positive for TUBB3 (86%). In contrast, the ODIn-PCSK9 neurons in the same batch, generated only 58% TUBB3-positive cells, see Figure 4.11, subfigure (A). However, the acquired number of cells was very low across both cell lines, which may contribute to confounding effects on the statistical significance of the acquired results. For the ODIn-PCSK9 cells, only 32 events in the TUBB3-positive subpopulation could be recorded, and for the ODIn-VEGFA2 neurons, the equivalent number of events was merely 216. Meanwhile, the neuroblastoma cell line SH-SY5Y was assayed as a positive control for the expression of TUBB3 and tested close to 100% positive for this marker (99.9%). For additional figures with SH-Sy5Y- and IgG-control conditions, see Figure D.6 in *Appendix D*, Section D.1. Note finally that the flow cytometry analysis of the second batch of ODIn-hiPSC-derived neurons is forthcoming and hence excluded from the scope of this thesis project.

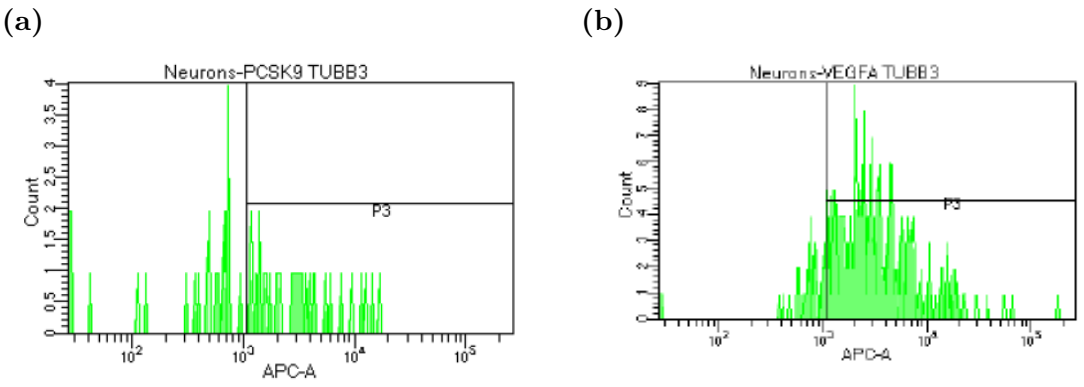


Figure 4.11: Flow cytometry analysis results of the first batch of ODIn-hiPSC-derived neurons stained with the APC-conjugated beta-tubulin class III (TUBB3), (a) TUBB3 stained ODIn-PCSK9, and (b) TUBB3 stained ODIn-VEGFA2. This batch of ODIn neurons was not stained for the pluripotency marker SSEA4, due to low cell count at harvest. One biological replicate per stain was assayed and the subpopulation of TUBB3-positive cells was gated based on forward- and side scatter (FSC/SSC), removing any cell duplets and smaller particles.

4.2 Targeted deep sequencing assays

Editing outcomes (indels and SNPs) at on- and off-target sites in the ODIn-hiPSC-derived cell types were determined by Amplicon-seq (AmpSeq) based targeted deep sequencing assays and analyzed with the AZ NGS analysis pipeline.

4.2.1 On-target editing validation with AmpSeq

On-target editing efficiency and checking for editing "leakage" in non-dox induced control samples was assessed in cell types before proceeding with the rhAmpSeq assay. A. Madsen verified that all generated ODIn-hiPS cell lines (PCSK9, VEGFA2

and HEK4) showed over 80% editing efficiency, as measured with targeted deep sequencing (AmpSeq) after induction with 1 µg/mL doxycycline for 48 hours. Furthermore, these cell lines exhibited low leakage-induced editing (1.8% for ODIn-PCSK9, 1.7% for ODIn-VEGFA2 and 6.1% for ODIn-HEK4) (Figure 4.12).

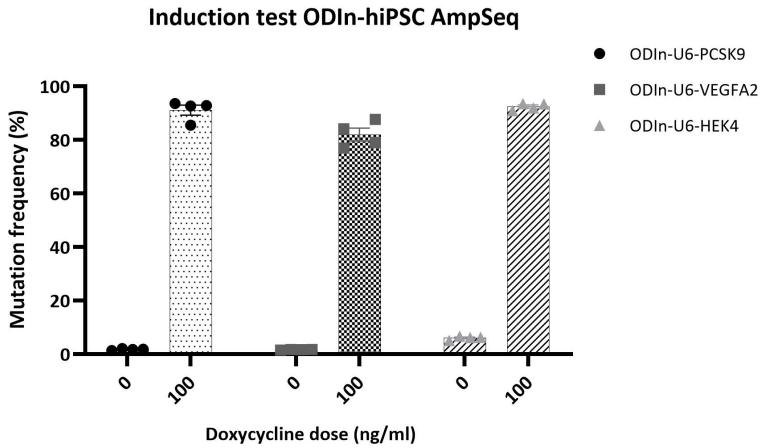
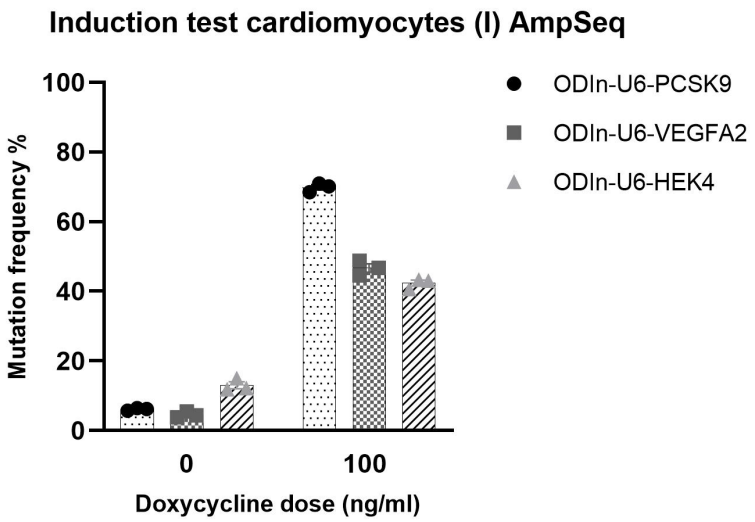


Figure 4.12: Dox-induction test with ODIn-hiPSCs to test the TET-ON inducible gene editing strategy

Cardiomyocytes

For the cardiomyocytes, overall, induced editing efficiencies ranged from 40.6-82.3%, while leakage expression generated 3.8-15.1% editing in non-induced cardiomyocytes, see Figure 4.13. on-target editing levels for dox-induced ODIn-PCSK9 samples were on par with the editing efficiencies observed in dox-induced ODIn-hiPSCs (68.5-81.3%). They were slightly higher in the second batch (71.1-81.3%), compared to the first batch (68.5-80.0%). However, leakage-induced editing also increased from batch 1 (5.7-6.4%) to batch 2 (9.7-10.5%), possibly indicating a correlation between non-induced leakage-mediated editing and induced editing (Figure 4.13). For the two cell lines with promiscuous guides (U6-VEGFA and U6-HEK4), on-target editing efficiencies were significantly lower, 44.5-48.8% for batch I and 40.4-48.7% in batch II of ODIn-VEGFA2 CMs, respectively. ODIn-HEK4 generated similar editing levels in batch I (40.6-43,4%) but slightly increased levels in batch II (51.5-54.8%). Meanwhile, ODIn-VEGFA2 presented the overall lowest leakage-induced on-target editing in untreated samples (3.8-5.5%), while ODIn-HEK4 presented the highest leakage (10.0-15.0%), across both batches.

(a)



(b)

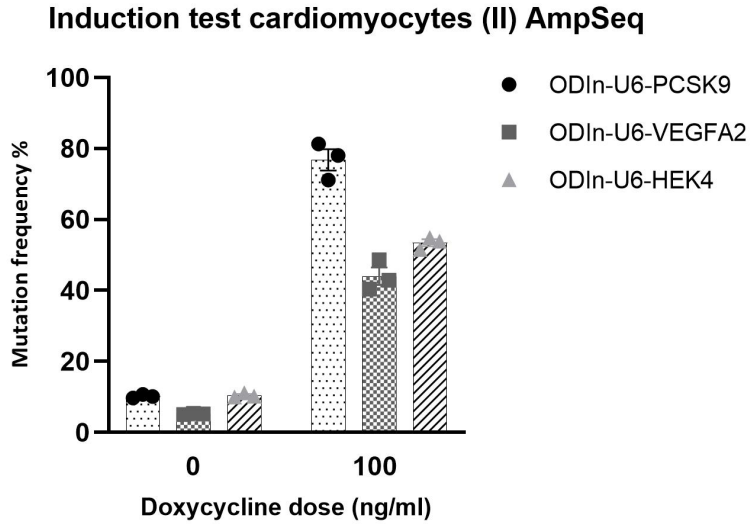


Figure 4.13: On-target editing efficiencies in control- and dox-induced (100 ng/mL for 48h) ODIn cardiomyocytes, respectively. (a) batch I and (b) batch II. Dotted bars w. circles: ODIn-PCSK9 CMs (*PCSK9 locus)*, chess-patterned bars w. squares: ODIn-VEGFA2 CMs (*VEGFA2* locus), and striped bars w. triangles: ODIn-HEK4 CMs (*HEK4* locus).

For the hepatocytes derived with the Takara Bio-protocol, targeted deep sequencing of the *PCSK9* and *HEK4* loci showed that induced editing efficiencies ranged between 57.0-60.6% for the ODIn-PCSK9 samples and 73.7-78.6% for the ODIn-HEK4 samples. Notably, this was on par with non-induced leakage editing levels, 51.7-56.2%, and 58.9-64.9%, respectively (Figure 4.14). This indicated that doxycycline or any doxycycline derivative could have been present in one or several of the Cellartis®differentiation media and thus induced both control samples and dox-treated samples prematurely. An investigation of the contents of the Takara Bio hepatocyte diff. kit was initiated, which resulted in the confirmed presence of a doxycycline derivative in one of the kit components. It was therefore decided to exclude this batch of hepatocytes from the rhampSeq assay. The novel STEMCELL Tech. kit for hepatic diff. was discovered and confirmed to not contain dox, and was therefore tested by A. Madsen for generating a new batch of hepatocytes. This batch was not matured in time for performing targeted sequencing analysis for on/off-target editing profiling and was therefore excluded from both AmpSeq and rhAmpSeq assays in this thesis project.

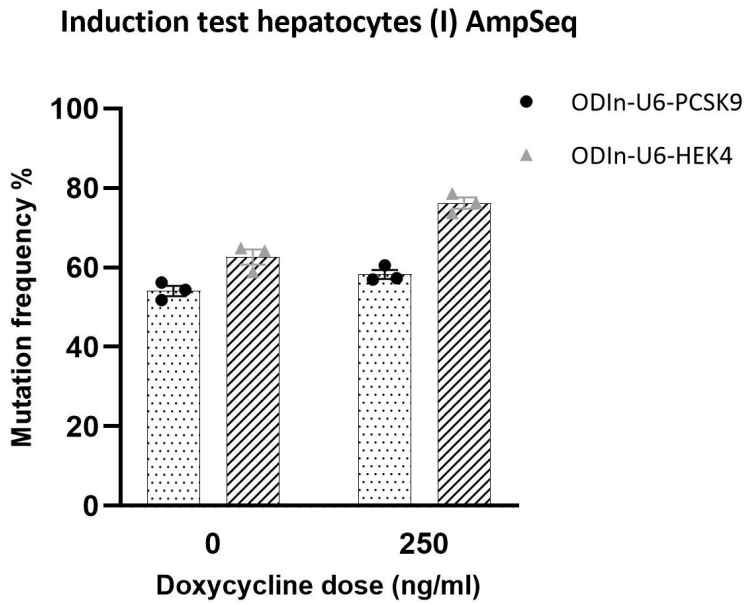


Figure 4.14: On-target editing efficiencies ODIn hepatocytes, batch I. Control (left) and dox-induced (250 ng/mL for 48h) heps (right) generated with the Takara Bio-protocol. Dotted bars w. circles: ODIn-PCSK9 hep (*PCSK9 locus)*, and striped bars w. triangles: ODIn-HEK4 hep (*HEK4* locus).

Targeted deep sequencing of the *PCSK9*- and *VEGFA2* loci in the first batch of ODIn neurons revealed significantly lower induced editing efficiencies compared to the other cell types. For ODIn-PCSK9 induced editing levels reached 13.3-16.8%, while ODIn-VEGFA2 showed slightly higher induced editing at 13.6-15.0%, see Figure 4.15. Meanwhile, non-induced leakage was comparatively high, 3.5-9.3% for the ODIn-VEGFA2 specimens and 5.0-6.4% for the ODIn-VEGFA2. Overall, on-target editing was approximately two-fold higher than non-induced leakage editing levels.

This comparatively small difference in on-target editing levels between conditions was considered a significant confounding effect in the downstream rhAmpSeq off-target analysis due to the sensitivity limit of the Illumina®NGS platform. Still, it was decided to include the neuronal samples from batch I in the final rhAmpSeq analysis, as the second batch would not be mature in due time of the NGS deadline. The results of rhAmpSeq off-target analysis on batch II neurons are forthcoming but outside the scope of this study.

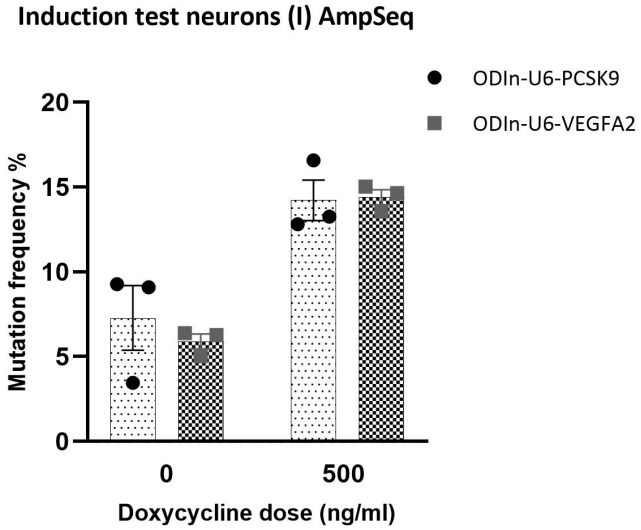


Figure 4.15: On-target editing efficiencies in ODIn neurons. Control samples (left) and dox-treated (500 ng/mL for 72h) samples (right) of the first batch of ODIn neurons (ODIn-PCSK9 and ODIn-VEGFA2). Dotted bars w. circles: ODIn-PCSK9 nr (*PCSK9 locus)*, chess-patterned bars w. squares: ODIn-VEGFA2 nr (*VEGFA2* locus).

4.2.1.1 Genotyping of ODIn neurons

To assess whether the two ODIn-hiPSC-derived neuron cell lines had been mixed up, somewhere during differentiation, or in fact, were mixed cultures, genotyping was performed. Three different assays were tested for this purpose; PCR, TIDE, and AmpSeq. PCR and TIDE-analysis were inconclusive. AmpSeq confirmed that the ODIn-VEGFA2 neuronal culture was, in fact, a mixed culture of ODIn-VEGFA2 and ODIn-PCSK9 neurons, but not vice versa. See *Appendix D*, Figure D.8 for an illustration of the identified induced- and leakage-mediated editing at both target loci in the mixed neuronal culture. For simplicity, the mixed culture will henceforth still be referred to as the ODIn-VEGFA2 neurons, but keep in mind that this is, in fact, a mixed culture. A thorough investigation of where during cell line generation- and differentiation this accidental mix-up occurred, was subsequently carried out. Targeted deep sequencing of both the *PCSK9*- and *VEGFA2* loci in both cell lines at the hiPSC-stage, NPC stage (after the third passage of neuronal induction), and mature neuronal cell stage, revealed that the mixup occurred somewhere during neuronal induction (before the third passage).

4.2.2 Off-target editing analysis with rhAmpSeq

In total, the on-target site and 92-152 highly ranked off-target (OT) sites, as determined in-house by CHANGE-seq, and cross-validated with CasOFFinder predictions and GUIDE-seq/CIRCLE-seq publications, were sequenced with rhAmpSeq targeted deep sequencing and analyzed with the AZ NGS analysis pipeline. Genomic DNA extracted from three different SpCas9-inducible ODIn cell types (hiPSC, cardiomyocytes and neurons) expressing either a specific guide RNA (PCSK9) or an unspecific/promiscuous guide (U6-VEGFA2/U6-HEK4), and either untreated or treated with doxycycline, were assayed for pursuing the aim of investigating potential cell-type effects on CRISPR-Cas9 off-target editing. More specifically this is the presence of significant editing events, i.e. the sum of small indels & SNPs with editing efficiency *geq* 0.1%, in different cell types induced with dox. For simplicity, a binary classification system of the filtered editing outcomes (0 = no editing & 1 = editing) was generated for visualization of the rhampSeq results. For further details on the rhampSeq analysis method, see Section 3.7.2.3 in *Methods*. Significant editing outcomes at on- and off-target sites were finally depicted in cell type-specific heatmaps, in separate figures for each CRISPR RGN, see Figures 4.16, 4.17, and 4.18.

4.2.2.1 rhAmpSeq results - PCSK9 panel

In Figure 4.16 the rhAmpSeq analysis results for each dox-treated sample expressing the specific RNA-guided nuclease (RGN) PCSK9-SpCas9, are depicted. The rhAmpSeq results are presented in three separate, cell type-specific heatmaps for facilitating comparison of the editing outcomes between cell types (hiPSCs, cardiomyocytes, and neurons). Notably, all of the submitted samples and target sites could be sequenced and mapped to a reference genome (GRCh38 - hg38). Furthermore, 88 out of 92 target sites (including *PCSK9*) passed the quality control and were included in the final assessment of potential cell type effects (Figure 4.16). Overall, on-target editing (*PCSK9* locus) could be verified also with rhAmpSeq, in all dox-treated cell types, except for one neuronal replicate. Moreover, off-target activity was low, as expected for the specific RGN. Low off-target activity is here defined as few significant editing events (*geq* 0.1% editing efficiency), as indicated by the low number of black cells across the three heatmaps. Nevertheless, some off-target activity could be detected, but only for a subset of replicates at each OT site. For the cardiomyocytes, one significant editing event was detected at off-target site 82 (OT82), which is in an unannotated region (non-protein-coding). For the neurons, editing was detected at OT02 (3/3 repl. mut. in neurons) and at OT75 (1/3 repl. mut. in hiPSCs). These are more interesting, as they appeared in more than one biological repl. per site. OT02 is a site within chromosome 1 open reading frame 198 (C1orf198) and OT75 is a site within the *phospholysine phosphohistidine inorganic pyrophosphate phosphatase* gene (LHPP). Finally, for the hiPSC specimens, nucleolytic activity was confirmed at OT36 (2/4 repl.) and at OT82 (2/4 repl.). OT36 is, like OT82, not annotated. Overall, comparing the nucleolytic activity of PCSK9-SpCas9 across all three cell types there are three off-target sites that differ significantly between cell types. These are OT02, OT036, and OT075, and of these

OT02 is the most interesting as it is present in all four of the neuronal replicates, the other two effects were only present in half of the assayed replicates ((Figure 4.16)). The identified cell type-specific off-target edits, especially at OT02, could be evidence of a significant cell type influence on off-target activity with specific CRISPR RGNs.

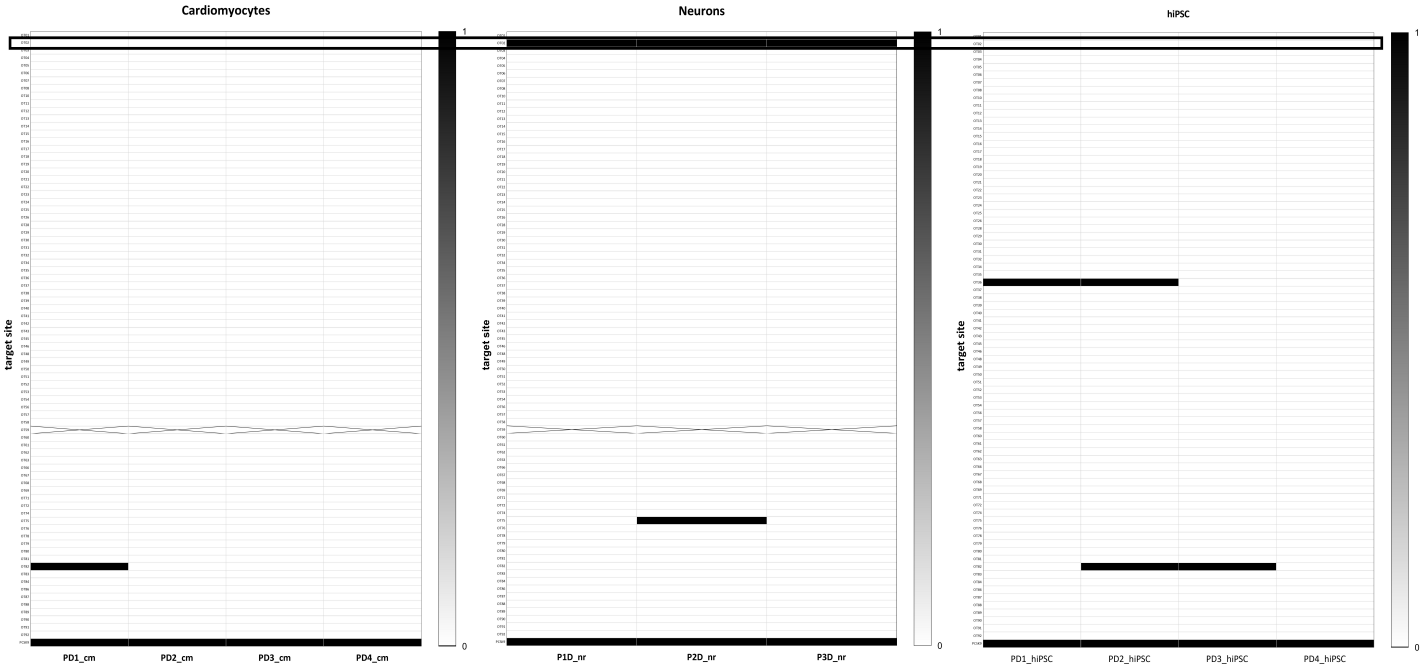


Figure 4.16: Assessment of on/off-target effects induced by the specific RGN PCSK9-SpCas9 by rhAmpSeq. The resulting on/off-target editing profiles in dox-treated ODIn-PCSK9 cell types are here presented in separate binary heatmaps. Cardiomyocytes (left), neurons (center), and hiPSCs (right). Genomic target sites are presented on the vertical axis (rows) and replicate samples on the horizontal axis (columns). In each heatmap, the cells represent a site- and sample-specific editing event, i.e., the total editing efficiency of small indels/SNPs. A significant editing event ($\geq$ 0.1% tot. editing efficiency) at a certain on- (*PCSK9*) or off-target site (*OTX*), and for a given sample, is highlighted in black. Contrarily, the absence of a significant editing event is indicated by white space. Crossed-over cells are genomic sites that did not pass the initial filtering analysis after sequencing. The most interesting editing events, where a certain mutation is present in all replicates of an exclusive cell type, are highlighted with a dashed box.

4.2.2.2 rhAmpSeq results - VEGFA2 panel

In Figure 4.17 the rhAmpSeq analysis results for each dox-treated sample expressing the unspecific RGN VEGFA2-SpCas9, are depicted. The rhAmpSeq results are presented in two separate heatmaps for facilitating comparison of the editing outcomes between cell types (cardiomyocytes, and neurons). Notably, the dox-treated replicates of one of the cell types, the ODIn-VEGFA2 hiPSCs, did not pass the

initial NGS pipeline. That is, these specimens produced very few mapped reads, and neither with any indication of introduced mutations. Thus, it was decided to exclude the ODIn-VEGFA2 hiPSCs entirely from subsequent analysis. As for the on-target editing, VEGFA-site 2 did not pass the initial quality control due to poor read count (≤ 500) in all replicates/cell types.

The two remaining cell types, cardiomyocytes (left in Figure 4.17) and neurons (right), displayed higher off-target activity, compared to the ODIn-PCSK9 samples, as expected for the unspecific RGNs. This is indicated by the increased number of black cells in the VEGFA2 heatmaps. Out of 152 off-target sites sequenced, 103 sites passed the QC/filtering and of these, 7 sites showed significant editing (black cells in Figure 4.17) across all replicates and cell types (site OT018, OT040, OT064, OT100, OT136, OT139, and OT145). The significant editing effects that differed between the two cell types were at site OT013 (2/4 repl. mut. in neurons), OT048 (1/4 repl. mut. in CMs) OT75 (3/4 repl. mut. in CMs), site OT107 (2/4 repl. mut. in CMs), OT115 (1/4 repl. mut. in neurons) and OT148 (3/4 repl. mut. in CMs). Most interesting are OT075, OT107, and OT148, which showed significant editing in 3 out of 4 biological replicates of cardiomyocytes, but not in neurons. OT75 maps to the annotated gene *acylphosphatase 2* (ACYP2), and OT148 is part of a non-coding RNA called *TSKU antisense RNA 1* (TSKU-AS1). In summary, VEGFA2-SpCas9 produced two potential off-target effects that differ between cell types in at least three biological replicates (OT075 and OT148). These identified cell type-specific off-target edits could be evidence of a significant cell type influence on off-target activity with unspecific CRISPR RGNs.

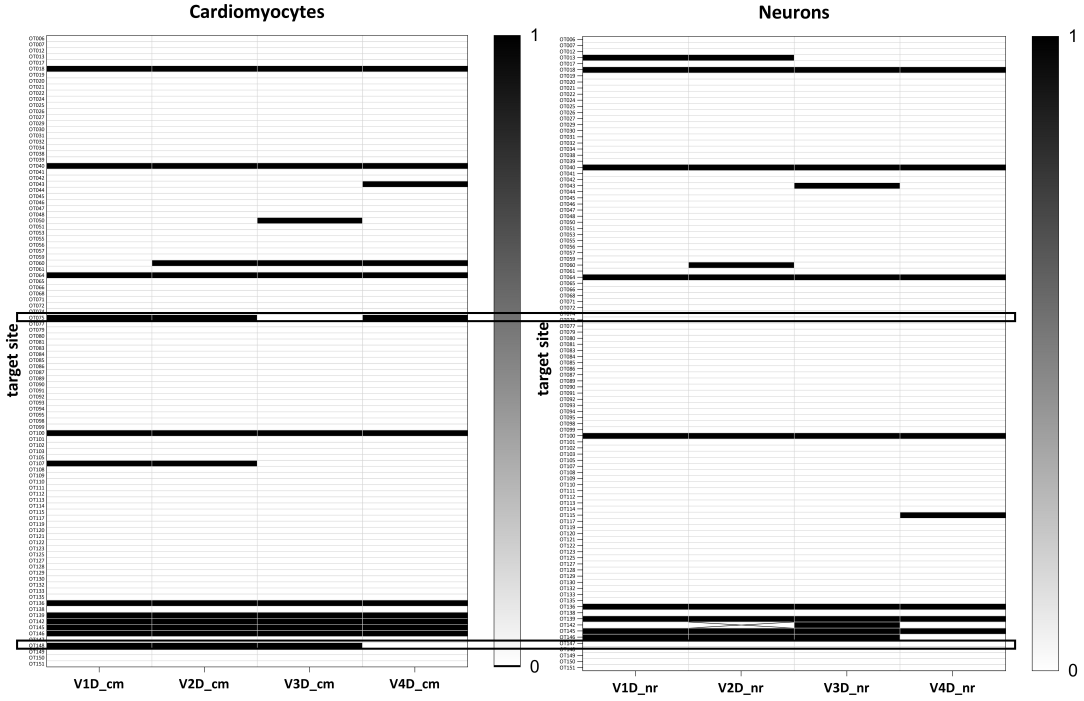


Figure 4.17: Assessment of on/off-target effects induced by the unspecific RGN VEGFA2-SpCas9 by rhAmpSeq. The resulting on/off-target editing profiles in dox-treated ODIn-VEGFA2 cell types are here presented in separate binary heatmaps. Cardiomyocytes (left), and neurons (right). Genomic target sites are presented on the vertical axis (rows) and replicate samples on the horizontal axis (columns). In each heatmap, the cells represent a site- and sample-specific editing event, i.e., the total editing efficiency of small indels/SNPs. A significant editing event ($\geq 0.1\%$ tot. editing efficiency) at a certain on- (*VEGFA2*) or off-target site (*OTX*), and for a given sample, is highlighted in black. Contrarily, the absence of a significant editing event is indicated by white space. Crossed-over cells are genomic sites that did not pass the initial filtering analysis after sequencing. The most interesting editing events, where a certain mutation is present in all replicates of an exclusive cell type, are highlighted with a dashed box.

4.2.2.3 rhAmpSeq results - HEK4 samples

Figure 4.18 presents the editing outcomes for each dox-treated sample expressing the promiscuous RGN HEK4-SpCas9 in separate heatmaps, cardiomyocytes (left in Figure 4.18) and hiPSCs (right). Editing outcomes in the ODIn-HEK4 cardiomyocytes were confounded by the lack of more than two biological replicates per site, as only two replicates could be sequenced. Two replicates were lost during rhAmpSeq library prep due to low gDNA output after PCR2. Hence, nucleolytic activity in remaining replicates cannot confidently be considered significant, but will anyways be regarded as potential true mutations. Nevertheless, all of the submitted samples could be sequenced and mapped to the reference genome.

On-target editing was verified in all dox-treated cell types. Furthermore, 112 out of the 120 off-target sites sequenced passed the quality control/filtering and were assessed for cell type variations. Overall, the HEK4-SpCas9 RGN presented the highest off-target activity compared to the other two RGNs. More specifically, 8 sites showed significant editing (black cells in Figure 4.18) in all replicates across both cell types; sites OT003, OT010, OT025, OT026, OT043, OT049, OT101, and OT114. The cardiomyocytes showed significant mutations at sites OT010, OT025, OT026, and O0T43, albeit in only one of the two replicate samples. The significant editing effects that differed between the two cell types were at sites OT005, OT008, and OT027 (3/3 repl. mut. in hiPSCs), OT015, OT022, OT034, OT068, OT093, OT095, OT102, OT115 and OT117 (1/3 repl. mut. in hiPSC), OT104 (1/2 repl. mut. in CMs), OT035 and OT055 (2/3 repl. mut. in hiPSCs). Of these, OT005, OT008, and OT027 are especially interesting, as they were edited in all biological replicates in either the cardiomyocytes or the hiPSCs, but not the other cell type. OT005 is located within the gene *centrosomal protein 89* (CEP89), OT008 maps to the gene *regulator of microtubule dynamics 3* (RMDN3) and OT027 is a site within the gene *interleukin 1 receptor accessory protein-like 2* (IL1RAPL2). Nevertheless, OT035 was mutated in two out of three biological replicates of hiPSCs and could therefore also be a potential cell-type effect (Figure 4.18). To summarize, HEK4-SpCas9 also demonstrated cell-type-specific mutations at three off-target sites, with all replicates per cell type confirming editing at each site. The identified cell type-specific off-target effects could provide further evidence of a significant cell type influence on off-target activity with unspecific CRISPR RGNs.

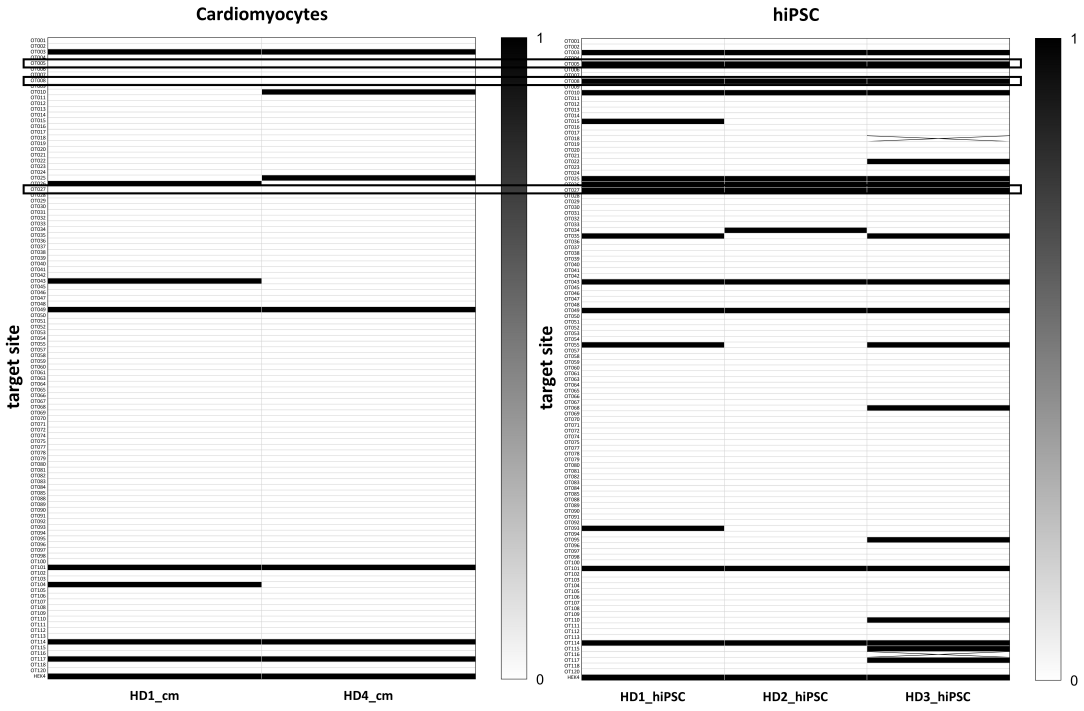


Figure 4.18: Assessment of on/off-target effects induced by the unspecific RGN HEK4-SpCas9 by rhAmpSeq. The resulting on/off-target editing profiles in dox-treated ODIn-HEK4 cell types are here presented in separate binary heatmaps. Cardiomyocytes (left), and hiPSCs (right). Genomic target sites are presented on the vertical axis (rows) and replicate samples on the horizontal axis (columns). In each heatmap, the cells represent a site- and sample-specific editing event, i.e., the total editing efficiency of small indels/SNPs. A significant editing event ($\geq$ 0.1% tot. editing efficiency) at a certain on- (*HEK4*) or off-target site (*OTX*), and for a given sample, is highlighted in black. Contrarily, the absence of a significant editing event is indicated by white space. Crossed-over cells are genomic sites that did not pass the initial filtering analysis after sequencing. The most interesting editing events, where a certain mutation is present in all replicates of an exclusive cell type, are highlighted with a dashed box.

5

Discussion

5.1 Interpretation of results & limitations

This thesis project builds upon the dual approach for OTE identification and validation, established by the VIVO study [2]. In this study, they established a combined approach for guide-specific *in vitro* screening of off-target cleavage sites and off-target editing validation *in vivo* in mice. Limitations of the study included analyzing OTE only in bulk tissue, using a mouse model instead of human cells, and focusing on OTEs in the target organ without considering such effects in peripheral tissues. In contrast, this thesis project rather aimed at investigating OTE mechanisms exclusively *in vitro*, in clinically relevant human cell types, by using hiPSC-derived cell lineages as a model for primary cell types. Moreover, the overall objective was to establish and validate an *in vitro*-based experimental pipeline within the Genome Engineering Department at AstraZeneca, for investigating such effects in clinically relevant cell types. Here, an optimized OTE identification and validation approach was successfully established by cross-validating the *in vitro*-identified (CHANGE-seq) off-target cleavage sites with independent methods (CasOFFinder/GUIDE-seq/CIRCLE-seq). Furthermore, off-target editing was in this thesis project identified for both promiscuous RGNs (VEGFA2-SpCas9/HEK4-SpCas9) and the specific RGN (PCSK9-SpCas9). In contrast, no off-target effects were identified for the mouse- or human PCSK9-SpCas9 in the VIVO-study [2], despite having used a similar targeted sequencing approach for OTE validation as in this thesis project. However, considerable differences in the study design between each respective study must be accounted for when making such a comparison.

Despite the improvements and novel objectives that were successfully explored in this thesis project, there are several limitations that hopefully can stimulate future improvements and continued research. The first part of this study focused on evaluating the pluripotency of the three established AZ-proprietary ODIn-hiPSC lines and using directed differentiation to generate valid experimental models for pursuing the stated research aim. To this end, five separate differentiation assays were performed and evaluated using functional and semi-quantitative cell characterization methods. The chosen cell types, three different human-induced pluripotent stem cell (hiPSC)-derived cell lineages with different cell characteristics, represent each of the three germ cell layers: a mid-brain dopaminergic neuronal cell lineage (ectoderm layer), a ventricular cardiomyocyte cell lineage (mesoderm layer), and a hepatocyte cell lineage (endoderm layer). The choice of cell types for the directed differentiation assays and downstream off-target analysis was based on the research

aim of investigating cell type effects on OTE. Cell types from different germ layers were considered a suitable strategy as germ layer formation occurs early on in natural embryogenesis and therefore creates phenotypically distinct cell types. Also, the choice was based on the availability of rigorously evaluated and high-quality commercial hiPSC-differentiation kits, without doxycycline-like antibiotics. Although this latter requirement was only confirmed beforehand by STEMCELL Tech. for the STEMdiff™kits and not by Takara Bio for the Cellartis®hepatocyte kit.

Considering the outcome of the cell characterization assays (microscopy, qPCR and FC), it is evident that the directed differentiation assays were overall successful. However, it is also clear that the generated cell populations possessed a significant level of heterogeneity, across all cell types. The microscopic examination of the cell morphology for each cell type provided a positive indication for successful differentiation, in all three cell types, batches, and replicate samples. All three ODIn-hiPSC lines presented typical hiPSC-like characteristics when compared to the provided reference microscopy images by Takara Bio (DEF-CS System User Manual). The differentiated cell types also presented morphological features in line with published references. However, some variability in morphological features and cell survival was also evident. Notably, the ODIn-HEK4 cardiomyocytes presented overall less beating compared to the other two cell lines, and only one of the promiscuous ODIn cell lines survived until the final stage of differentiation during neuron diff. (ODIn-VEGFA2) and Cellartis®hepatocytes diff. (ODIn-HEK4). A plausible explanation for these findings is the potential cytotoxic effects imposed by leakage-mediated off-target editing. If for example, non-induced off-target nucleolytic activity occurs in genomic regions involved in cell growth and/or cell viability, or even yet, cell differentiation, such activity could explain the observed issues with the promiscuous cell lines, where OTE was expected (and later verified) to be higher than in the cultures with the specific PCSK9-SpCas9.

Comparative gene expression was measured with qPCR to validate the different cell types on a population level. qPCR showed an overall upregulation of the cell-specific markers in all three cell lines, indicating that, on average, the hiPSCs were pluripotent and this was later confirmed by the successful development into differentiated cell types. Nevertheless, variability between replicates was high in all cell lines, and the relative upregulation (fold change) of lineage-specific markers was not as high as in comparative literature [7, 14, 30, 37, 60, 108]. Flow cytometry analysis also demonstrated a strong cell-cell heterogeneity. In fact, only a low proportion of differentiated cell types reached the maturation stage according to the low proportion of cells that stained positive for the cell-specific markers (2-86% across cell types). Only the neurons presented expression levels on par with the kit manufacturer's specifications (58-86%), yet these results were considerably confounded by a low cell count at acquisition. That is, only 32 events could be recorded for the ODIn-PCSK9 and 216 events for the ODIn-VEGFA neurons (approximately 10000 is recommended by BD Biosciences). However, when taking into account the expression of pluripotency markers (POU5F1 and SSEA4), the results are somewhat contradictory. The qPCR assays suggested strong downregulation of the POU5F1 marker, suggesting reduced pluripotency. Together with the upregulation of cell-type specific markers, this indicated successful differentiation towards the different

cell types. Yet, immunostaining with the SSEA4 marker indicated the opposite, a significant proportion of cells in each lineage and cell line were in fact still pluripotent (approximately 20-30% of stained cells tested positive). It should be said that the qPCR results provide limited confidence about the success of each differentiation assay as the depicted gene expression levels only represent overall sample means and therefore omits subpopulation heterogeneity. Therefore, it can be argued that a valid experimental model for our research aim could not be produced and that a majority of the cells in the final cell populations were in fact progenitor cells rather than mature cell types. Still, after having repeated the differentiation assays for all three cell types, with adjustments to the seeding densities (cardiomyocytes and neurons), the kit manufacturer (hepatocytes), cell dissociation protocol (cardiomyocytes), and prolonged maturation time, all three protocols were successfully optimized and a larger proportion of mature cell types obtained.

The second part of this study was dedicated to first validating on-target editing with AmpSeq and then investigating off-target editing effects in the hiPSC-derived cell lineages with rhAmpSeq. On-target editing was verified in all three cell lineages and for each guide RNA in the surviving cell lines, although with varying induced editing efficiencies (20-80% for hiPSCs, cardiomyocytes, and neurons) and significant non-induced leakage-mediated editing in each cell type (1-20% for hiPSCs, cardiomyocytes, and neurons). Some batch variations in induced editing efficiencies and leakage-mediated editing were also identified by AmpSeq. These variations may partly be attributed to technical issues during cell harvest and AmpSeq library preparation, as well as the implemented changes between different batches during differentiation (seeding density, dox-dose, allowed time for DNA-repair after dox-induction etc.). They may also partly be attributed to inherent biological differences between- and within the generated (heterogeneous) cell types. It has for example been reported that the cell type, and cell cycle stage, influence the pathway, efficiency, and kinetics of DNA repair after CRISPR-mediated double-strand breaks (DSBs) [38, 53, 56]. For example, it has been shown that the endogenous DNA repair machinery is slower in neurons [68], which may explain why the induced editing efficiency in the ODIn neurons was lower than in the other cell types.

The choice of promoters for the CRISPR-Cas9 constructs may also be a limiting factor in this study. Although the TET-ON system has been used extensively for controlled expression of transgenes [34, 35, 66], including transient expression of Cas9 nucleases, there are also demonstrated limitations. For example, as previously reported [35] and further demonstrated in this study, TET-ON promoters exhibit some basal expression in the absence of an inducer (here doxycycline), leading to background expression of the target gene even when the system is intended to be inactive (leakage). Also, induction of gene expression with TET-ON promoters has been shown to demonstrate variable kinetics and efficiency, that is, the time it takes for the inducer (e.g., doxycycline) to reach a certain concentration and activate gene expression can vary, as well as the timing and extent of induction between different cell types and experimental conditions [35]. A delayed or too short induction of SpCas9 expression may prove insufficient for RGN assembly and subsequent gene editing. Furthermore, doxycycline has been previously shown [112], and in this study verified, to exhibit cytotoxic potential. Cytotoxic potential is a

broad term but it is here defined as the ability of the inducing agent to cause reduced cell viability at higher concentrations and/or after prolonged exposure. In this study, dox-mediated cytotoxicity was observed in trilineage differentiated and dox-treated ODIn cells at a concentration of 100 ng/mL and after 48h of incubation. This cytotoxicity was further demonstrated in dox-induced undifferentiated ODIn-VEGFA2 hiPSCs, but not in control specimens, the other two hiPSC lines, or in mature cell types. Therefore, it can be hypothesized that in induced undifferentiated hiPSCs with promiscuous RGNs, there might be a combined effect of dox-mediated chemical cytotoxicity and induced off-target editing resulting in a loss of functional genes for hiPSC proliferation and survival. However, as extensive cell death was also observed in the trilineage-differentiated and dox-induced ODIn-PCSK9 cells, the cytotoxic properties of doxycycline must still be considered an independent factor. It can moreover be hypothesized that mitotic- and differentiating cells are more sensitive to doxycycline and that this sensitivity may diminish with the loss of pluripotency or due to other mechanisms involved in differentiation. Besides the cytotoxic effects of doxycycline, the aspect of induction efficiency must also be considered and further tailored to the intended cell type for improving targeted editing efficiency and reducing leakage-mediated editing.

Note that the dox-dose regiment was established based on a combination of internal recommendations from experienced researchers and externally published literature. However, the recommendations did not reach a consensus on either dose, incubation time, or the optimal time for subsequent DNA repair, for either of the cell types. For example, the suggested dox-dose for inducing TET-ON controlled CRISPR systems in neurons ranged from 100-2000 ng/ml in published literature, with incubation times ranging from 24h to 72h and with limited provided information on subsequent maintenance time for DNA repair [3, 92, 87]. As such, the dox induction assay can be considered a significant contributor to sub-optimal induced on-target editing efficiency. Especially as the neurons were exposed the longest (72h) to dox and thereby the CRISPR nuclease, which prolongs the window for nucleolytic activity at both on- and off-target sites.

There are also a few reported limitations associated with the use of U6-promoters for sgRNA expression, which is another limiting factor of this study. For once, the U6 promoter has been reported to require a specific sequence context for efficient initiation of transcription, which may limit the selection of target sites for sgRNA-integration [54]. Also, the strength of the U6 promoter has been reported to vary depending on the specific sgRNA sequence and the genomic context at the integration site [32]. This variability in expression levels can affect the efficiency of Cas9-mediated cleavage and subsequent genome editing, leading to inconsistent outcomes across different target sites. Thus, to verify gRNA expression, a gRNA-specific qPCR assay is forthcoming.

The influence of cell type on CRISPR-Cas9 off-target editing (OTE) was here investigated using rhAmpSeq. Each RGN showed 1-3 cell type-specific off-target mutations that were present in all induced replicates, and several other OTE that were present in a subset of replicates, for an individual cell type. As expected, the specific PCSK9-SpCas9 system generated the lowest number of significant off-target mutations, after filtering out low-quality sites and normalization against leakage-

mediated editing in control samples. Meanwhile, the HEK4-SpCas9 introduced the highest number of significant lineage-specific OTE. This provides further evidence that the specificity of the CRISPR RNA correlates with off-target nucleolytic activity, yet, that off-target editing occurs also for specific RGNs. Furthermore, the significant off-target activity could explain the loss of the ODIn-HEK4 cell line in neuronal differentiation as well as the significantly less mature cardiomyocyte culture obtained in both batches. Note that despite the positive indication of cell-type effects on CRISPR off-target editing, the rhAmpSeq results were also confounded by replicate variability, i.e. the presence of mutations only in a subset of replicates.

Moreover, background noise in the rhAmpSeq results was high. That is, the presence of seemingly "significant" ($\geq 0.1\%$) mutations at on- and off-target sites in non-induced control samples. This background noise can be attributed to two main factors, 1. leakage-mediated editing in control samples, and 2. sequencing bias. The former issue was confirmed by the previous on-target editing validation assays (AmpSeq), and was highly cell type-dependent. Therefore, the downstream off-target analysis was inevitably biased towards the cell types with less leakage-mediated editing compared to on-target editing efficiency. The second factor, sequencing bias, is attributed to the sequencing method used (short-read Illumina®NGS) and can impact the accuracy and completeness of the sequencing results. For example by introducing false positive- or false negative mutations. However, with the paired-end sequencing approach used, potential sequencing bias can be mitigated to some extent. Nevertheless, these confounding effects may have resulted in an inaccurate calling of actual off-target mutations and a biased assessment of cell type off-target effects.

The observed variations in induced editing efficiencies and overall high leakage-mediated editing limits the overall confidence in accurately evaluating cell type-specific CRISPR off-target effects with the rhAmpSeq assay. Especially also given the sensitivity limit of the applied targeted sequencing method (short-read Illumina®NGS). As such, less frequent edits ($\leq 0.1\%$) in samples with high leakage-mediated editing might have been filtered out in the subsequent rhAmpSeq assay and thus potential cell-type specific edits could have been missed, or alternatively overestimated in other cell types. It is for example likely that the off-target activity in the first batch of ODIn neurons was underestimated, as the reported difference between induced editing and leakage-mediated editing was merely two-fold (see Figure 4.15). This issue emphasizes the need for improved mutation detection sensitivity. Nevertheless, the rhAmpSeq assay confirmed that promiscuous guides generate more edits than the specific guide, but challenges with low-frequency and low-read count edits obstructed the accurate distinguishing of truly absent mutations from false negatives. Furthermore, the off-target editing analysis method employed to investigate cell-type effects on CRISPR OTE was semi-biased, that is, some potential off-target edits at unidentified loci might have been missed. I.e. instead of validation through whole-genome sequencing which looks at the entire genome for off-target effects, this strategy employs targeted deep sequencing to verify editing at predicted on- and off-target sites in each cell type. Moreover, the online version of CasOFFinder which was here used to cross-validate the identified (CHANGE-seq) off-target cleavage sites for the PCSK9-guide is limited to output only a subset of

all the predicted genomic cleavage sites, therefore it cannot be excluded that some significant OT sites for this guide may have been overlooked.

The observed leakage-mediated on-target editing detected by AmpSeq in mature cell types, but not at hiPSC-stage, may be attributed to epigenetic modifications and chromatin remodeling during differentiation. For example, a previous study [21] found that differentiation-mediated chromatin remodeling can activate previously heterochromatic genomic regions. This activation could hypothetically also stimulate CRISPR nucleolytic activity at on- and off-target sites in the absence of doxycycline and thereby cause leakage-mediated mutations. Furthermore, as chromatin structure and accessibility vary naturally among different cell types and this influences the exposure of both on- and off-target cleavage sites to the CRISPR system and other proteins involved in subsequent DNA repair processes, this can also be a contributory factor to the observed differences in induced editing efficiencies between cell types. These effects must however be investigated thoroughly in the established ODIn cell types before drawing any further conclusions.

Lastly, as verified by the AmpSeq genotyping assay with the ODIn neurons, the final ODIn-VEGFA2 culture was heterogeneous in terms of genotype, as it was actually a mixed culture of both ODIn-VEGFA2 cells and ODIn-PCSK9 cells. This has inevitable confounding effects on downstream rhAmpSeq off-target editing analysis. For once, the validated off-target edits in the ODIn-VEGFA2 neuronal specimens are actually a combination of cleaved targets in both cell lines, i.e. induced by two different RGNs. This means that the unspecific cleavage by the promiscuous VEGFA2-guide may be overestimated when only looking at the rhAmpSeq results in the mixed culture. However, as the U6-PCSK9 guide is target specific, it is less likely to present off-target activity, especially at genomic sites with high sequence homology to the U6-VEGFA2 target sequence, as this differs significantly from the U6-PCSK9 target sequence (see Table B.1, in *Appendix B*). This was indeed observed in the rhAmpSeq results for the ODIn-PCSK9-specimens which showed an overall low prevalence of significant off-target mutations. It can furthermore be argued that the reverse mix-up would be more deleterious for the validation of off-target effects with the "specific" guide, as it would likely be confounded by U6-VEGFA2-mediated cleavage at the PCSK9-specific off-target sites and thus overestimate this RGN's off-target nucleolytic activity. To verify this hypothesis, a comparison of sequence homology between the identified U6-VEGFA2 off-target cleavage sites (CHANGE-seq/GUIDE-seq and CIRCLE-seq results) and the off-target sites identified for the U6-PCSK9 (CHANGE-seq/CasOFFinder) could provide an estimate for the likelihood of false-positive mutation calling in such a mixed culture.

5.2 Outlook

As mentioned in the *Results*, the flow cytometry analysis of the second batch of ODIn neurons, with optimized cell density, is forthcoming. These results will provide further insight into the population homogeneity and with that validation of the success of neuronal differentiation. Furthermore, a second rhAmpSeq off-target analysis assay is planned for the near future and will include a new batch of ODIn-hiPSC, potentially the first batch of cardiomyocyte (for comparison to the second batch OTE profiles), the second batch of ODIn hepatocytes (STEMCELL Tech. kit) and the second batch of ODIn-neurons. For this assay, thorough troubleshooting of the rhampSeq library preparation protocol will be performed and optimization strategies implemented. Aspects to consider here are the initial dilution of gDNA which has to be evaluated in relation to the sample elution reagent used (Qiagen EB or Lucigen QE), as this may affect subsequent amplification steps. Moreover, the number of amplification cycles applied during both PCR 1 and PCR 2 may need to be further optimized to increase the throughput of samples from library prep to NGS analysis. Also, the issue with poor on-target sequencing of the loci targeted by the promiscuous guides must be addressed, with emphasis on the *VEGFA2* locus, as the rhAmpSeq assay presented an overall low coverage of this site.

Farther into the future it can additionally be of interest to repeat the study with significant improvements to the gene editing strategy, hiPS cell line generation, downstream directed differentiation, and doxycycline induction assays. One of the most significant confounding effects in the assessment of cell-type effects on OTE in this study is the leakage-mediated editing that was verified (to different extents) in all cell types. Therefore, it can be argued that the chosen gene editing strategy must be revised, and novel engineering strategies employed for controlling the expression and assembly of the CRISPR RGN more closely. This includes withstanding external factors that may cause non-induced editing, such as endogenous transcription factors and differentiation-associated chromatin remodeling. A promising strategy for this pursuit is using an engineered degron-fused Cas9 nuclease [58] that would enable the selective elimination of Cas9 from the cell upon leakage-mediated expression. An important aspect to consider here is how to avoid degron-mediated Cas9 degradation after dox-induced expression, to ensure sufficient time for induced nucleolytic activity. Having optimized the CRISPR-Cas9 strategy, the next step would be to redo the directed differentiation assays to achieve more homogenous cell populations in the final cell types, to reduce the distorting effects of cell heterogeneity on the final OTE analysis and interpretation of results in the context off cell type effects.

An approach to optimize the inducible CRISPR-Cas strategy would be to explore other promoters than the doxycycline-inducible TET-ON system for SpCas9 expression and the U6-promoter for sgRNA expression. Moreover, a way to monitor leakage-mediated editing and mitigate downstream effects on the OTE analysis could be to continuously measure GFP expression in the cells throughout hiPSC culture and differentiation assays, as the integrated SpCas9-cassette in the *AAVS1* locus also contains a GFP reporter gene [66]. Continuous GFP monitoring can subsequently be performed either by harvesting of cells for immunocytochemical char-

acterization with an anti-GFP antibody or with live cell imaging by fluorescence microscopy. If GFP is expressed before induction, then termination of the affected cells or colony picking/cell sorting of GFP-negative cells should be considered before proceeding to dox-induction.

To optimize the differentiation assays, several factors can be considered, including the choice of cell type(s) and differentiation protocol, as well as the experimental design of the differentiation assays. In this study, the choice of cell type for the directed differentiation was based on both the objective to generate phenotypically distinguished cell types and using available differentiation kits due to time sensitivity issues. However, other cell types from the same germ layers could be considered with other complete differentiation kits or by the cocktail approach of preparing all the necessary differentiation media *in-house*. To improve overall culture homogeneity in the differentiated cell types, cell-sorting methods should be considered. As an example, Fluorescence-Activated Cell Sorting (FACS) based on GFP expression after dox-induction could be used to isolate a homogenous population of mature cells expressing the CRISPR RGN. However, this requires additional strategies for preventing co-sorting of false-positives, in terms of progenitor cells with leakage-induced GFP expression. Another issue with sorting strategies is obtaining a sufficient viable cell count after sorting for ensuring a high enough DNA/RNA yield for subsequent analyses.

Having optimized the cell line generation and differentiation assays, the next step could be to optimize the cell characterization assays for better validation of the generated cell types. If performed properly, without technical issues, the combination of morphological assessment, qPCR, and flow cytometry analysis would provide both qualitative and quantitative evidence for successful differentiation on both the cell population level and single-cell level. Yet, within the frame of this study, where technical issues at multiple stages before- and during the characterization assays occurred, poor performance in terms of low expression of cell type markers can be misinterpreted as true biological effects. Therefore, a more rigorous literature review into strategies for minimizing technical bias should be performed, as well as including additional markers for each of the ODIn cell types, e.g. testing other pluripotency markers than POU5F1 (and NANOG), and other lineage-specific markers.

It should also be noted that although iPSC-derived cell types are valuable experimental models for studying the potential cell type effects on CRISPR-Cas9 off-target editing, these effects must be validated in more relevant cell types for the intended downstream application. In the context of therapeutic genome editing, this study needs to be repeated using primary cell types, preferably derived from both healthy- and relevant patient populations. Also, these effects must be validated in both target- and peripheral tissues of more complex- and clinically relevant *in vivo* models such as mice and non-human primates. The latter objective is already in focus of an ongoing study within the same research team at AZ.

An important parallel optimization approach for the detection of rare off-target effects ($\leq 0.1\%$), besides reducing leakage-mediated editing, is to employ novel sequencing approaches with improved sensitivity and accuracy, such as single-cell sequencing or long-read sequencing technologies. This is especially important for clinically relevant (specific) RGNs where conventional Illumina®NGS-based meth-

ods fall short. Promising novel sequencing methods like DUPLEX-seq [88] could potentially be used to further characterize less frequent mutations, but currently, it is considered too cost-inefficient for use in these applications. Moreover, to characterize the off-target effects induced in the different cell types, including identifying translocations and large deletions, additional assays are needed. For example, based on the generated rhAmpSeq data, an *in silico* tool like CRISPECTOR [5] could be employed for identifying translocation events between on- and off-target sites for each guide RNA in the dox-induced samples. Then, PCR amplification techniques, such as digital-droplet PCR, could be employed to validate these translocations or large deletions [111].

As replicate variability and background noise from control samples can introduce challenges in accurately assessing off-target editing, moving forwards, it is important to carefully design experimental controls and replicate experiments to minimize these effects and ensure reliable and reproducible results. For example, including DNA from an independent cell line that does not express any gRNA and will therefore not be edited at all, would provide a control for rhAmpSeq to filter out any false-positive OT sites. Furthermore, as the Illumina®-based rhAmpSeq assay holds limitations, it is important to integrate multiple complementary assays and validation methods, such as *in silico* predictions with more comprehensive datasets and advanced machine learning algorithms, to improve overall confidence in the OT analysis results. Also, improving the established functional assays, like the hiPSC trilineage differentiation kit, could help validate and confirm experimental designs at an early research stage and thereby reduce lead times and inefficient resource use in OTE research.

5.3 Implications for therapeutic applications

The aim of this MSc Thesis project was to expand on our understanding of the mechanisms behind CRISPR-Cas9 off-target editing, as a prerequisite for developing safer human gene therapies. Despite variable experimental outcomes, this project provided new insights into the mechanisms underlying CRISPR-Cas9-mediated off-target editing. It identified potential cell type-specific effects that occur independently of nuclease specificity. Once validated and further investigated, these cell type effects on CRISPR-Cas-mediated off-target editing must be taken into account when targeting different cell types and tissues for therapeutic purposes. Minimizing off-target activity in genome editing enhances efficiency, precision, and safety in therapeutic applications.

To improve safety and efficacy, advanced machine learning algorithms can be developed to consider cell-specific genomic context, such as chromatin structure and epigenetic modifications, for target site evaluation and selection. Avoiding high-risk regions with sequence homology to the guide sequence in multiple validated cell types can further enhance safety and efficacy. However, it should be noted that while in silico prediction tools and guide design are valuable, they cannot perfectly predict actual off-target cleavage. Therefore, they alone are insufficient to avoid CRISPR-mediated off-target editing.

Understanding cell type effects can also inform the development of more efficient delivery methods for CRISPR components. Targeted delivery systems can minimize exposure of non-target cells and peripheral tissues to the CRISPR system, reducing the likelihood of off-target editing in those cell types. Furthermore, raising awareness of off-target effects can guide regulatory agencies and policymakers in formulating guidelines and regulations for the clinical development and commercialization of therapeutic genome editing technologies.

To ensure the safe and effective treatment of patients with CRISPR-Cas-based therapies, rigorous risk assessments and monitoring of off-target effects should be integrated into the national and international regulatory framework. This approach promotes the responsible and ethical use of CRISPR-based therapies throughout the drug development pipeline and clinical application phases.

6
Conclusions

In this thesis project, an experimental investigation was conducted to explore potential cell-type effects on CRISPR-SpCas9 off-target gene editing in human iPSC-derived cell lineages. Three different *in house*-established (AZ) isogenic and recombinant human induced pluripotent stem cell (hiPSC) lines with integrated TET-ON-inducible CRISPR-SpCas9 constructs were used to derive the experimental models. Three different hiPSC-derived cell lineages; ventricular cardiomyocytes, hepatocytes, and midbrain dopaminergic neurons, were generated and partly validated, through directed differentiation and cell characterization assays. Transient gene editing was induced through the addition of doxycycline in mature cell types, and on/off-target editing was subsequently analyzed using the rhAmpSeq targeted deep sequencing assay. rhAmpSeq identified potentially significant cell-specific off-target editing effects for each guide RNA, that were moreover independent of the CRISPR nuclease specificity. Further optimization of the differentiation assays and subsequent rhampSeq library preparation is required for confirming these cell type-specific mutations, and additional assays are recommended for characterizing these effects in depth. Nevertheless, the rhAmpSeq results indicate that the cell type indeed influences the CRISPR-Cas9-mediated off-target activity and these findings could have considerable implications for future clinical applications of CRISPR genome editors to combat genetic diseases.

Bibliography

[1] A Adan et al. “Flow cytometry: basic principles and applications”. In: *Critical reviews in biotechnology* 37.2 (2017), pp. 163–176.

[2] P Akcakaya et al. “In vivo CRISPR editing with no detectable genome-wide off-target mutations”. In: *Nature* 561.7723 (2018), pp. 416–419.

[3] AA Akhtar et al. “Inducible expression of GDNF in transplanted iPSC-derived neural progenitor cells”. In: *Stem Cell Reports* 10.6 (2018), pp. 1696–1704.

[4] A Al Abbar et al. “Induced pluripotent stem cells: Reprogramming platforms and applications in cell replacement therapy”. In: *BioResearch Open Access* 9.1 (2020), pp. 121–136.

[5] I Amit et al. “CRISPECTOR provides accurate estimation of genome editing translocation and off-target activity from comparative NGS data”. In: *Nature Communications* 12.1 (2021), p. 3042.

[6] S Bae, J Park, and JS Kim. “Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases”. In: *Bioinformatics* 30.10 (2014), pp. 1473–1475.

[7] N Balafkan et al. “A method for differentiating human induced pluripotent stem cells toward functional cardiomyocytes in 96-well microplates”. In: *Scientific reports* 10.1 (2020), pp. 1–14.

[8] M Ballester et al. “Direct conversion of human fibroblast to hepatocytes using a single inducible polycistronic vector”. In: *Stem Cell Research & Therapy* 10 (2019), pp. 1–10.

[9] M Bär, D Bär, and B Lehmann. “Selection and validation of candidate housekeeping genes for studies of human keratinocytes—review and recommendations”. In: *Journal of Investigative Dermatology* 129.3 (2009), pp. 535–537.

[10] E Bianconi et al. “An estimation of the number of cells in the human body”. In: *Annals of human biology* 40.6 (2013), pp. 463–471.

[11] MF Bolukbasi, A Gupta, and SA Wolfe. “Creating and evaluating accurate CRISPR-Cas9 scalpels for genomic surgery”. In: *Nature methods* 13.1 (2016), pp. 41–50.

[12] SJ Brouns et al. “Small CRISPR RNAs guide antiviral defense in prokaryotes”. In: *Science* 321.5891 (2008), pp. 960–964.

[13] S Cancellieri et al. “Human genetic diversity alters off-target outcomes of therapeutic gene editing”. In: *Nature Genetics* 55.1 (2023), pp. 34–43.

[14] A Carpentier et al. “Hepatic differentiation of human pluripotent stem cells in miniaturized format suitable for high-throughput screen”. In: *Stem Cell Research* 16.3 (2016), pp. 640–650.

[15] A Cebrian-Serrano and B Davies. “CRISPR-Cas orthologues and variants: optimizing the repertoire, specificity and delivery of genome engineering tools”. In: *Mammalian Genome* 28 (2017), pp. 247–261.

[16] SW Cho et al. “Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases”. In: *Genome research* 24.1 (2014), pp. 132–141.

[17] L Cong et al. “Multiplex genome engineering using CRISPR/Cas systems”. In: *Science* 339.6121 (2013), pp. 819–823.

[18] DBT Cox, RJ Platt, and F Zhang. “Therapeutic genome editing: prospects and challenges”. In: *Nature medicine* 21.2 (2015), pp. 121–131.

[19] TJ Cradick et al. “CRISPR/Cas9 systems targeting β-globin and CCR5 genes have substantial off-target activity”. In: *Nucleic acids research* 41.20 (2013), pp. 9584–9592.

[20] E Deltcheva et al. “CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III”. In: *Nature* 471.7340 (2011), pp. 602–607.

[21] JR Dixon et al. “Topological domains in mammalian genomes identified by analysis of chromatin interactions”. In: *Nature* 485.7398 (2012), pp. 376–380.

[22] SH Doak and ZM Zaır. “Real-time reverse-transcription polymerase chain reaction: technical considerations for gene expression analysis”. In: (2012), pp. 251–270.

[23] JA Doudna. “The promise and challenge of therapeutic genome editing”. In: *Nature* 578.7794 (2020), pp. 229–236.

[24] LE Dow. “Modeling disease in vivo with CRISPR/Cas9”. In: *Trends in molecular medicine* 21.10 (2015), pp. 609–621.

[25] G Education. “Cell culture basics handbook”. In: *Thermo Fisher Scientific* 541 (2016).

[26] S Författningssamling. “Lag (2003: 460) om etikprövning av forskning som avser människor”. In: *Stockholm: Utbildningsdepartementet* (2003).

[27] H Frangoul et al. “CRISPR-Cas9 gene editing for sickle cell disease and β-thalassemia”. In: *New England Journal of Medicine* 384.3 (2021), pp. 252–260.

[28] RL Frock et al. “Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases”. In: *Nature biotechnology* 33.2 (2015), pp. 179–186.

[29] Y Fu et al. “High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells”. In: *Nature biotechnology* 31.9 (2013), pp. 822–826.

[30] X Gao et al. “Hepatocyte-like cells derived from human induced pluripotent stem cells using small molecules: implications of a transcriptomic study”. In: *Stem Cell Research & Therapy* 11 (2020), pp. 1–21.

[31] JE Garneau et al. “The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA”. In: *Nature* 468.7320 (2010), pp. 67–71.

[32] LA Gilbert et al. “Genome-scale CRISPR-mediated control of gene repression and activation”. In: *Cell* 159.3 (2014), pp. 647–661.

[33] JD Gillmore et al. “CRISPR-Cas9 in vivo gene editing for transthyretin amyloidosis”. In: *New England Journal of Medicine* 385.6 (2021), pp. 493–502.

[34] F González et al. “An iCRISPR platform for rapid, multiplexable, and inducible genome editing in human pluripotent stem cells”. In: *Cell stem cell* 15.2 (2014), pp. 215–226.

[35] M Gossen and H Bujard. “Tight control of gene expression in mammalian cells by tetracycline-responsive promoters.” In: *Proceedings of the National Academy of Sciences* 89.12 (1992), pp. 5547–5551.

[36] C Graham and S Hart. “CRISPR/Cas9 gene editing therapies for cystic fibrosis”. In: *Expert Opinion on Biological Therapy* 21.6 (2021), pp. 767–780.

[37] T Grancharova et al. “A comprehensive analysis of gene expression changes in a high replicate and open-source dataset of differentiating hiPSC-derived cardiomyocytes”. In: *Scientific Reports* 11.1 (2021), pp. 1–21.

[38] A Gupta et al. “Targeted chromosomal deletions and inversions in zebrafish”. In: *Genome research* 23.6 (2013), pp. 1008–1017.

[39] DE Handy, R Castro, and J Loscalzo. “Epigenetic modifications: basic mechanisms and role in cardiovascular disease”. In: *Circulation* 123.19 (2011), pp. 2145–2156.

[40] ND Heintzman et al. “Histone modifications at human enhancers reflect global cell-type-specific gene expression”. In: *Nature* 459.7243 (2009), pp. 108–112.

[41] P Hermans et al. “Insertion element IS987 from Mycobacterium bovis BCG is located in a hot-spot integration region for insertion elements in Mycobacterium tuberculosis complex strains”. In: *Infection and immunity* 59.8 (1991), pp. 2695–2705.

[42] D Hockemeyer and R Jaenisch. “Induced pluripotent stem cells meet genome editing”. In: *Cell stem cell* 18.5 (2016), pp. 573–586.

[43] I Höijer et al. “CRISPR-Cas9 induces large structural variants at on-target and off-target sites in vivo that segregate across generations”. In: *Nature Communications* 13.1 (2022), p. 627.

[44] S Hrvatin et al. “Differentiated human stem cells resemble fetal, not adult, β cells”. In: *Proceedings of the National Academy of Sciences* 111.8 (2014), pp. 3038–3043.

[45] PD Hsu et al. “DNA targeting specificity of RNA-guided Cas9 nucleases”. In: *Nature biotechnology* 31.9 (2013), pp. 827–832.

[46] W Hu et al. “RNA-directed gene editing specifically eradicates latent and prevents new HIV-1 infection”. In: *Proceedings of the National Academy of Sciences* 111.31 (2014), pp. 11461–11466.

[47] X Hu et al. “The application and progression of CRISPR/Cas9 technology in ophthalmological diseases”. In: *Eye* 37.4 (2023), pp. 607–617.

[48] I Illumina. *Amplicon Sequencing Solutions*. `https://www.illumina.com/techniques/sequencing/dna-sequencing/targeted-resequencing/amplicon-sequencing.html`. 2023.

[49] G Inak et al. “Concise review: induced pluripotent stem cell-based drug discovery for mitochondrial disease”. In: *Stem Cells* 35.7 (2017), pp. 1655–1662.

[50] Integrated DNA Technologies. *Integrated DNA Technologies*. `https://eu.idtdna.com/pages/products/crispr-genome-editing/rhampseq-crispr-analysis-system?utm_source=google`. 2023.

[51] Y Ishino et al. “Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product”. In: *Journal of bacteriology* 169.12 (1987), pp. 5429–5433.

[52] R Jansen et al. “Identification of genes that are associated with DNA repeats in prokaryotes”. In: *Molecular microbiology* 43.6 (2002), pp. 1565–1575.

[53] N Javaid and S Choi. “CRISPR/Cas system and factors affecting its precision and efficiency”. In: *Frontiers in Cell and Developmental Biology* 9 (2021), p. 761709.

[54] M Jinek et al. “A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity”. In: *science* 337.6096 (2012), pp. 816–821.

[55] M Jinek et al. “Structures of Cas9 endonucleases reveal RNA-mediated conformational activation”. In: *Science* 343.6176 (2014), p. 1247997.

[56] P Karran. “DNA double strand break repair in mammalian cells”. In: *Current opinion in genetics & development* 10.2 (2000), pp. 144–150.

[57] HR Kempton and LS Qi. “When genome editing goes off-target”. In: *Science* 364.6437 (2019), pp. 234–236.

[58] N Khajanchi and K Saha. “Controlling CRISPR with small molecule regulation for somatic cell genome editing”. In: *Molecular Therapy* 30.1 (2022), pp. 17–31.

[59] K Kim et al. “Epigenetic memory in induced pluripotent stem cells”. In: *Nature* 467.7313 (2010), pp. 285–290.

[60] YL Kuang et al. “Evaluation of commonly used ectoderm markers in iPSC trilineage differentiation”. In: *Stem cell research* 37 (2019), p. 101434.

[61] A Kyttälä et al. “Genetic variability overrides the impact of parental cell type and determines iPSC differentiation potential”. In: *Stem cell reports* 6.2 (2016), pp. 200–212.

[62] CR Lazzarotto et al. “CHANGE-seq reveals genetic and epigenetic effects on CRISPR–Cas9 genome-wide activity”. In: *Nature biotechnology* 38.11 (2020), pp. 1317–1327.

[63] S Lessard et al. “Human genetic variation alters CRISPR-Cas9 on-and off-targeting specificity at therapeutically implicated loci”. In: *Proceedings of the National Academy of Sciences* 114.52 (2017), E11257–E11266.

[64] S Li et al. “Universal toxin-based selection for precise genome engineering in human cells”. In: *Nature communications* 12.1 (2021), pp. 1–14.

[65] CA Lino et al. “Delivering CRISPR: a review of the challenges and approaches”. In: *Drug delivery* 25.1 (2018), pp. 1234–1257.

[66] A Lundin et al. “Development of an ObLiGaRe doxycycline inducible Cas9 system for pre-clinical cancer drug discovery”. In: *Nature communications* 11.1 (2020), pp. 1–16.

[67] EZ Macosko et al. “Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets”. In: *Cell* 161.5 (2015), pp. 1202–1214.

[68] R Madabhushi, L Pan, and LH Tsai. “DNA damage and its links to neurodegeneration”. In: *Neuron* 83.2 (2014), pp. 266–282.

[69] K Makarova et al. “An updated evolutionary classification of CRISPR-Cas systems”. In: *Nature reviews. Microbiology* (Sept. 2015). DOI: 10.1038/nrmicro3569.

[70] KS Makarova et al. “A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action”. In: *Biology direct* 1.1 (2006), pp. 1–26.

[71] P Mali et al. “CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering”. In: *Nature biotechnology* 31.9 (2013), pp. 833–838.

[72] LA Marraffini and EJ Sontheimer. “CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA”. In: *science* 322.5909 (2008), pp. 1843–1845.

[73] A Martinez-Fernandez et al. “iPS cell–derived cardiogenicity is hindered by sustained integration of reprogramming transgenes”. In: *Circulation: Cardiovascular Genetics* 7.5 (2014), pp. 667–676.

[74] E Metzl-Raz et al. “Generation of a dual edited human induced pluripotent stem cell Myl7-GFP reporter line with inducible CRISPRi/dCas9”. In: *Stem cell research* 61 (2022), p. 102754.

[75] FJ Mojica et al. “Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements”. In: *Journal of molecular evolution* 60 (2005), pp. 174–182.

[76] M Morrison et al. “The European General Data Protection Regulation: challenges and considerations for iPSC researchers and biobanks”. In: *Regenerative medicine* 12.6 (2017), pp. 693–703.

[77] J Muhr and KM Ackerman. “Embryology, gastrulation”. In: (2020).

[78] A Nakata, M Amemura, and K Makino. “Unusual nucleotide arrangement with repeated sequences in the Escherichia coli K-12 chromosome”. In: *Journal of bacteriology* 171.6 (1989), pp. 3553–3556.

[79] J O’Malley et al. “High-resolution analysis with novel cell-surface markers identifies routes to iPS cells”. In: *Nature* 499.7456 (2013), pp. 88–91.

[80] DG Ousterout et al. “Multiplex CRISPR/Cas9-based genome editing for correction of dystrophin mutations that cause Duchenne muscular dystrophy”. In: *Nature communications* 6.1 (2015), p. 6244.

[81] Y Panina et al. “Validation of common housekeeping genes as reference for qPCR gene expression analysis during iPS reprogramming process”. In: *Scientific reports* 8.1 (2018), p. 8716.

[82] SB Park et al. “A dual conditional CRISPR-Cas9 system to activate gene editing and reduce off-target effects in human stem cells”. In: *Molecular Therapy-Nucleic Acids* 28 (2022), pp. 656–669.

[83] D Paull et al. “Automated, high-throughput derivation, characterization and differentiation of induced pluripotent stem cells”. In: *Nature methods* 12.9 (2015), pp. 885–892.

[84] FA Ran et al. “Genome engineering using the CRISPR-Cas9 system”. In: *Nature protocols* 8.11 (2013), pp. 2281–2308.

[85] E Regulation. “679 of the european parliament and of the council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general

data protection regulation)". In: *EC (General Data Protection Regulation)* (2016).

[86] N Savić and G Schwank. "Advances in therapeutic CRISPR/Cas9 genome editing". In: *Translational Research* 168 (2016), pp. 15–21.

[87] B Schmid et al. "Generation of two gene edited iPSC-lines carrying a DOX-inducible NGN2 expression cassette with and without GFP in the AAVS1 locus". In: *Stem Cell Research* 52 (2021), p. 102240.

[88] MW Schmitt et al. "Detection of ultra-rare mutations by next-generation sequencing". In: *Proceedings of the National Academy of Sciences* 109.36 (2012), pp. 14508–14513.

[89] M Schubert et al. *Evaluate CRISPR-Cas9 edits quickly and accurately with rhAmpSeq targeted sequencing.*

[90] T Seki et al. "Generation and characterization of functional cardiomyocytes derived from human T cell-derived induced pluripotent stem cells". In: *PLoS One* 9.1 (2014), e85645.

[91] K Sekine et al. "Robust detection of undifferentiated iPSC among differentiated cells". In: *Scientific reports* 10.1 (2020), p. 10293.

[92] R Sheta et al. "Combining NGN2 programming and dopaminergic patterning for a rapid and efficient generation of hiPSC-derived midbrain neurons". In: *Scientific Reports* 12.1 (2022), p. 17176.

[93] JW Shin et al. "Permanent inactivation of Huntington's disease mutation by personalized allele-specific CRISPR/Cas9". In: *Human molecular genetics* 25.20 (2016), pp. 4566–4576.

[94] S Shin et al. "Comprehensive analysis of genomic safe harbors as target sites for stable expression of the heterologous gene in HEK293 cells". In: *ACS Synthetic Biology* 9.6 (2020), pp. 1263–1269.

[95] Y Smurnyy et al. "DNA sequencing and CRISPR-Cas9 gene editing for target validation in mammalian cells". In: *Nature chemical biology* 10.8 (2014), pp. 623–625.

[96] B Song et al. "Improved hematopoietic differentiation efficiency of gene-corrected beta-thalassemia induced pluripotent stem cells by CRISPR/Cas9 system". In: *Stem cells and development* 24.9 (2015), pp. 1053–1065.

[97] V Tabar and L Studer. "Pluripotent stem cells in regenerative medicine: challenges and recent progress". In: *Nature Reviews Genetics* 15.2 (2014), pp. 82–92.

[98] V Tadić et al. "CRISPR/Cas9-based epigenome editing: An overview of dCas9-based tools with special emphasis on off-target activity". In: *Methods* 164 (2019), pp. 109–119.

[99] K Takahashi and S Yamanaka. "Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors". In: *cell* 126.4 (2006), pp. 663–676.

[100] K Takahashi et al. "Induction of pluripotent stem cells from adult human fibroblasts by defined factors". In: *cell* 131.5 (2007), pp. 861–872.

[101] *The molecular scissors correcting defective genes.* URL: `https://www.astrazeneca.com/r-d/next-generation-therapeutics/therapeutic-gene-editing.html`. (accessed: 26.09.2022).

[102] SQ Tsai and JK Joung. "Defining and improving the genome-wide specificities of CRISPR–Cas9 nucleases". In: *Nature Reviews Genetics* 17.5 (2016), pp. 300–312.

[103] SQ Tsai et al. "GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases". In: *Nature biotechnology* 33.2 (2015), pp. 187–197.

[104] SQ Tsai et al. "High-fidelity CRISPR-Cas9 variants with undetectable genome-wide off-targets". In: *Nature* 529.7587 (2016), pp. 490–495.

[105] SQ Tsai et al. "CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR–Cas9 nuclease off-targets". In: *Nature methods* 14.6 (2017), pp. 607–614.

[106] J Tycko, VE Myer, and PD Hsu. "Methods for optimizing CRISPR-Cas9 genome editing specificity". In: *Molecular cell* 63.3 (2016), pp. 355–370.

[107] F Uddin, CM Rudin, and T Sen. "CRISPR gene therapy: applications, limitations, and implications for the future". In: *Frontiers in oncology* 10 (2020), p. 1387.

[108] GS Virdi et al. "Highly enriched hiPSC-derived midbrain dopaminergic neurons robustly models Parkinson's disease". In: *bioRxiv* (2020), pp. 2020–09.

[109] L Vossaert et al. "Reference loci for RT-qPCR analysis of differentiating human embryonic stem cells". In: *BMC molecular biology* 14 (2013), pp. 1–7.

[110] K Watanabe et al. "A ROCK inhibitor permits survival of dissociated human embryonic stem cells". In: *Nature biotechnology* 25.6 (2007), pp. 681–686.

[111] HL Watry et al. "Rapid, precise quantification of large DNA excisions and inversions by ddPCR". In: *Scientific reports* 10.1 (2020), pp. 1–11.

[112] J Xie, A Nair, and TW Hermiston. "A comparative study examining the cytotoxicity of inducible gene expression system ligands in different cell types". In: *Toxicology in vitro* 22.1 (2008), pp. 261–266.

[113] XH Zhang et al. "Off-target effects in CRISPR/Cas9-mediated genome engineering". In: *Molecular Therapy-Nucleic Acids* 4 (2015), e264.

A

Additional laboratory protocols and reagents

This appendix provides detailed, step-wise laboratory procedures for all experiments, as well as motivations and clarifications for the selected methods.

A.1 General cell culture & molecular biology protocols

In this section, some general cell culture- and molecular biology protocols used throughout this thesis project are disclosed. Important general principles for maintaining and differentiating hiPS cells and hiPS-derived cell lineages include using a sterile technique; always using PPE (gloves, shoe covers, and laboratory coat), wiping all surfaces, including gloves and reagents that might come into contact with cells with 70% ethanol, and perform all sterile work in a laminar-flow ventilated biosafety cabinet. Furthermore, optimal cell culture conditions must be ensured, by means of a suitable combination of nutrients, growth factors, hormones, buffers, and a temperature- and pressure-regulated incubator with fixed settings of O_2 and CO_2 gas-exchange, and proper (tissue-treated) culture vessels [25]. Below is a summary of necessary standard laboratory equipment for culturing the ODIn-hiPSC cells and downstream cell lineages in this study, but not included in the list of purchased reagents and equipment.

1. Biohazard safety cabinet certified for Level II handling of biological materials.
2. Incubator with humidity and gas control to maintain 37°C and 95% humidity in an atmosphere of 5% CO_2 in air.
3. Low-speed centrifuge (200-500g) with a swinging bucket rotor.
4. Pipette-aid with appropriate serological pipettes.
5. Vacuum pump with aspiration inlet and appropriate aspiration pipettes.
6. Pipettor with appropriate tips.
7. 15 mL and 50 mL conical tubes.
8. Inverted microscope with a total magnification of 10X to 100X.
9. Isopropanol freezing container.
10. -80°C freezer.
11. -20°C freezer.
12. Refrigerator ($2 - 8$°C).

A.1.1 Cell freezing protocol

The ODIn-hiPSCs and ODIn-hiPSC-derived cell lineages were frozen at different time points to preserve the cells for later assays. Here, a general freezing protocol for cryopreserving the different ODIn-cell types is presented. First, freezing media was prepared by diluting DMSO to a final concentration of 10% (v/v) in a suitable cell culture media, i.e FBS (hepatocyte progenitor cells and hepatocytes), specialized serum-free freezing medium for pluripotent cells (hiPSCs), or KO-SR (neuronal progenitor cells and neurons). 1 mL freezing agent and 1-4 million cells are sufficient for each cryotube but the exact cell count should be optimized for each cell type and downstream application. Cells were harvested by one of four protocols, see Section A.1.2 and collected in a conical tube. Cell count was determined with an automated cell counter (Vi-CELL XR Cell Viability Analyzer, Beckman Coulter) and as many cells as needed for each cryotube were transferred into a clean falcon tube. Cells were pelleted by centrifuging at 300g (hiPSCs) to 500g (NPCs/neurons) for 5 minutes at room temperature (RT). Subsequently, the supernatant was aspirated and the cell pellet was resuspended in the prepared freezing medium, to achieve the desired freezing concentration (e.g. 2 million cells per ml). Quickly, as DMSO is toxic to the cells at prolonged exposure in a liquid state, the cell suspension was transferred to cryotubes and placed in a Mr. Frosty for controlled freezing in a $-80°C$ freezer for at least 24h. Frozen cells were then stored at $-80°C$ for up to 6 months. For long-term storage liquid nitrogen was instead used.

A.1.2 Cell dissociation

Four different cell protocols were used for harvesting the different cell types at passage and for downstream analyses. A detailed protocol for each approach; TrypLE (cat. nr. 12604013, Gibco™), Gentle cell Dissociation Reagent ("GCDR", cat. nr. 100-0485, STEMCELL™Technologies), Accutase™(cat. nr. 07920, STEMCELL™Technologies), and STEMdiff™Cardiomyocyte Dissociation reagent ("SCDR", cat. nr.05025, STEMCELL™Technologies), is here provided together with an indication of which cell type they were applied for.

A.1.2.1 TrypLE™cell dissociation protocol

The TrypLE protocol was used for cell culture applications with the Cellartis®System (DEF-CS 500 System for ODIn-hiPSC culture and the Cellartis®Hepatocyte Differentiation) from Takara bio. First, all media were removed from the culture vessels and the cells were gently washed once with a suitable washing media, e.g. 0.1 $\mathrm{mL\,cm^{-2}}$ of PBS−/− or Dulbecco's Modified Eagle's Medium w. 15 mL Hepes ("DMEM/F-12", cat. nr. 36254, STEMCELL™Technologies) at room temperature. The washing agent was removed by aspiration or with a pipette and room temperature TrypLE Select 1x was carefully added to the wall of the tilted vessel, enough to cover the entire cell culture surface ($20\mu\mathrm{l/cm^2}$). The vessel was then incubated at 37°C for 3-5 minutes or until cells had visibly detached. Cell detachment was verified under a microscope and if cells were still attached, the side of the vessel was gently thumped against a hard surface (e.g. the side of a bench). When all cells had

detached, the plate was carefully tilted and the TrypLE was carefully discarded with a pipette, ensuring that no cells were removed with it. The cells were subsequently resuspended in $3x0.1$ mL cm^{-2} of a suitable culture medium and transferred to a conical tube for downstream applications. Cell aggregates were broken up as necessary by pipetting up and down. For more details see *Cellartis®DEF-CS Culture System User Manual*.

A.1.2.2 Gentle Cell Dissociation Reagent

The GCDR protocol was used for maintaining ODIn-hiPSCs in mTeSR™1/CMC and for the first STEMdiff™ventricular cardiomyocyte differentiation (STEMCELL™ Tech.). All media was removed from the culture vessel and the cells gently washed once with 0.1 mL cm^{-2} of a suitable washing media, e.g. PBS−/− or DMEM/F-12, warmed to room temperature. The washing agent was removed by aspiration or with a pipette and 0.1 mL cm^{-2} of GCDR at room temperature was carefully added to the wall of the tilted vessel. The cells were incubated at room temperature for 8-12 minutes or until they had visibly detached from the bottom. The GCDR was removed with a pipette and cells were subsequently resuspended in 3x 0.1 mL cm^{-2} of mTeSR™1, and transferred to a conical tube. Cell aggregates were broken up as necessary by pipetting up and down and then used for downstream applications. See *Enzyme-Free Passaging of Human Pluripotent Stem Cells Using Gentle Cell Dissociation Reagent* (STEMCELL™Technologies) for more details.

A.1.2.3 Accutase®

Accutase®(STEMCELL™Technologies) was used for harvesting the neuronal progenitor cells and mature neurons, derived with the STEMdiff™Midbrain Neuron Differentiation Kit (STEMCELL™Technologies). For harvesting cells with Accutase®, all of the media from the cell culture vessel was aspirated. As rinsing with PBS is not necessary, 0.1 mL cm^{-2} of prewarmed Accutase®(37°C) was immediately added to the vessel. Cells were incubated at room temperature for 5-10 minutes until the cells have rounded and no longer appear "spidery" (neuronal morphology). Without discarding the Accutase, as no neutralization steps are required, cells were dislodged and transferred to a conical tube prepared with 0.2 mL cm^{-2} of prewarmed DMEM/F-12 (STEMCELL™Technologies). The remaining cells were resuspended in 3x 0.1 mL cm^{-2} of warm DMEM/F-12 (STEMCELL™Technologies) and added to the conical tube.

A.1.2.4 STEMdiff™Cardiomyocyte Dissociation reagent

For the second cardiomyocyte differentiation assay with the STEMdiff™Ventricular Cardiomyocyte Differentiation Kit (STEMCELL™Technologies), the specialized SCDR protocol was used (STEMCELL™Technologies). The manufacturer's instructions were followed without alterations, see A.1.2.4

A.1.3 DNA/RNA extraction

Total genomic DNA was extracted by one of two approaches, either with QuickExtract™DNA Extraction Solution ('QE', cat. nr.QE0905T, QE09050, Lucigen® Epicenter®)or with one of the Qiagen DNA/RNA extraction protocols; RLT Plus Lysis buffer ('RLT', cat. nr. 79216, Qiagen) followed by combined DNA/RNA extraction using the Qiagen AllPrep®Mini- or Micro Kit (cat. nr. 80204; nr. 80284, Qiagen) or RNA purification exclusively using the RNeasy RNA Mini- or Micro Kits (cat. nr. 74104; nr. 74004, Qiagen). RNA was extracted exclusively with the latter of the two approaches.

A.1.3.1 QuickExtract (QE) DNA extraction

The former approach was taken for fast DNA isolation with downstream applications including genotyping of the iPSC-derived neuronal progenitor cells as well as for the dox-titration assay on ectodermal cells. The latter approach was taken when preparing for simultaneous purification of genomic DNA and total RNA, i.e. DNA for the on/off-target editing analysis and mRNA for the qPCR cell characterization assay. With the QE approach, fridge-cold QE (500 µL for one well of a 6-well plate or 50 µL for one well of a 96-well plate) was added directly to the cells after having removed the culture media. Cell membranes were disrupted by pipetting up-and-down x10 times and 3x50 µL cell lysate per well harvested was transferred to PCR tubes (three technical replicates per biological replicate/well). The harvested samples were subsequently heat treated to lyse the remaining cellular/ tissue material in a thermocycler at 70°C for 10 minutes, followed by 98°C for 10 min, and then cooled down to 4°C if not used immediately.

A.1.3.2 Qiagen DNA/RNA purification protocols

The same cell lysis and homogenization process was applied for all Qiagen AllPrep kits. Subsequent DNA/RNA purification steps varied for each kit (Qiagen AllPrep Mini/Micro). For ODIn-hiPSCs and ventricular cardiomyocytes cultured in 6-well and 12-well plates, the Qiagen DNA/RNA AllPrep Mini Kit (cat. nr. 80204, Qiagen) was used to achieve higher DNA/RNA yields. Hepatocytes and neurons, cultured in 96-well plates, were purified using the Qiagen AllPrep Micro kit (cat. nr. 80284, Qiagen).

Cells grown as monolayers in cell-culture vessels were either harvested by one of the cell dissociation protocols (see Section A.1.2) and collected as a cell pellet before lysis, or directly lysed in the vessel. iPSCs and neurons, being less sticky, were lysed directly, while ventricular cardiomyocytes and hepatocytes were first trypsinized and pelleted to ensure thorough removal of extracellular components that could compromise downstream nucleic acid purification. The sticky cells were harvested, pelleted at 300-500g for 5 minutes at room temperature (RT), and the supernatant completely removed before proceeding to cell lysis. For neurons and hiPSCs, the culture media was removed and cells were washed once with 0.1 mL cm^{-2} of PBS−/− before proceeding. Next, 350 µL/well of Qiagen RLT Lysis buffer (cat. nr. 79216, Qiagen) with 10 µL/mL added β-mercaptoethanol was added. Cells were lysed

by pipetting up and down 10 times, and the cell lysate was transferred to 1.5 mL Eppendorf tubes kept on ice. The cell lysates were homogenized for a minimum of 30 seconds by vortexing (3b). The homogenized cell lysates were either used directly for DNA/RNA extraction or kept at −20°C for a maximum of one week or at −80°C for longer storage before extraction. The homogenized lysate was transferred to an AllPrep DNA spin column placed in a 2 mL collection tube. The tubes were centrifuged for 30 seconds at ≥ 8000g (≥ 10,000 rpm).

The AllPrep DNA spin column, containing the DNA, was placed in a new 2 mL collection tube (supplied) and stored at room temperature (15 − 25°C) or at 2 − 8°C for later DNA purification. The flow-through was immediately used for RNA purification. For total RNA purification, one volume of 70% ethanol was added to the flow-through and mixed well by pipetting. Up to 700 µL per sample was transferred to an RNeasy (AllPrep Mini) or RNeasy MinElute (AllPrep Micro) spin column placed in a 2 mL collection tube. The lid was closed gently and the tube(s) centrifuged for 15 s at ≥ 8000g (≥ 10,000 rpm) in RT. The flow-through was completely discarded and sample(s) washed with 700 µL Buffer RW1 and centrifuged at ≥ 8000g (≥ 10,000 rpm), in RT. The flow-through was completely discarded and the membrane washed twice with 500 µL Buffer RPE (AllPrep Mini). For the AllPrep Micro protocol, the second RPE wash was substituted for 80% ethanol. Sample(s) were centrifuged for 15 s (wash I) and 2 mins (wash II) at ≥ 8000g (≥ 10,000 rpm), RT. Spin column(s) were placed in a new 2 mL collection tube and centrifuged at full speed for 1 min (AllPrep Mini). Or, with an opened lid, at full speed for 5 min (AllPrep Micro). The collection tube was then discarded with the flow-through and placed in a new 1.5 mL collection tube. 30 µL (AllPrep Mini) or 14 µL (AllPrep Micro) of RNase-free water was added directly to the spin column membrane and tubes centrifuged for 1 min at ≥ 8000g (≥ 10,000 rpm) to elute the RNA. Eluted RNA was placed on ice and transferred to a −80°C freezer until further use.

For genomic DNA purification the AllPrep DNA spin column(s) were washed once with 500 µL of Buffer AW1 and centrifuged for 15 s at ≥ 8000g (≥ 10,000 rpm) in RT. The flow-through was completely discarded and the spin column membrane(s) washed once with 500 µL Buffer AW2 by centrifuging for 2 min at full speed. The collection tube(s) were discarded and the spin column(s) placed in new 1.5 mL collection tube(s). Genomic DNA was extracted with 100 µL room-temperature Buffer EB (AllPrep Mini), or 50 µL of 70°C prewarmed Buffer EB (AllPrep Micro), added directly to the spin column membrane(s). Tube(s) were incubated at RT for 1 min and then centrifuged for 1 min at ≥ 8000g (≥ 10,000 rpm). Eluted genomic DNA were stored in a −20°C freezer until further use.

A.2 ODIn-hiPSC line generation

The following is a summary of the ODIn-hiPSC line generation based on the ObLiGaRe- and Xential-approaches [66, 64]. Cell line generation was not performed by the author.

A.2.1 Cell line and reagents

Cell line

Human iPSCs containing a dox-inducible SpCas9-cassette integrated into the genomic AAVS1-locus, *ODInCas9C86, p9* (vial obtained from Anders Lundin [66])

Reagents

1. Cellartis®DEF-CS™500 Kit, including media, growth factors, and coating (cat. nr. Y30010, Takara Bio)
2. TrypLE Select Enzyme (cat. nr.12563011, ThermoFisher)
3. PBS (−/−), no calcium (Ca^{2+}), no magnesium (Mg^{2+}) (cat. nr.14190250, ThermoFisher)
4. PBS (+/+), with calcium and magnesium (cat. nr.14040-091, Gibco™)
5. FuGENE HD (cat. nr. E2311, Promega)
6. Diphtheria toxin (Cat. nr. D0564-1MG, Sigma-Aldrich)

A.2.2 Human iPSC culture

Human iPSCs (*ODInCas9C86, p9*) were cultured using DEF-CS™pluripotent culturing system according to the manufacturer's instructions (*Cellartis®DEF-CS Culture System User Manual*).

A.2.3 Transfection

- Transfection of hiPSC was done at the cellular growth log phase using FuGENE HD reverse transfection. Transfection was performed in 48-well plates (1.1 cm^2) with 42,000 cells per well in 200 μL complete growth medium.
- Molar ratios of Cas9, guide, and HDR donor: 2.5:1:10 (66 ng Cas9 + 17 ng gRNA + 167 ng HDR template).
- FuGENE HD:DNA ratio of 1:1 is defined as 0.1 μL FuGENE HD to 100 ng DNA; use a ratio of 2.5:1.
- Total amount of transfected DNA per 48-well plate: 250 ng.
- For each sample, set up $n = 3$ technical replicates.

Day (−2)

- Harvest and seed cells to a confluency of 40,000cm^{-2} into 2x T25 flasks (1 x 10^6 cells per flask) before transfections in order to transfect at the growth log phase after 48 hours.

Day 0: Day of transfection

- Change media ∼6h before transfection to replenish growth factors and boost proliferation.

Cell transfections and diphtheria toxin selection

A 48-well TC-treated flat-bottom plate (Corning®Costar®) was coated with DEF-CS™Coat-1 (1:20 in D-PBS+/+), and incubated for at least 20 min in 37°C. Media was removed from cells and washed 1x with PBS−/−. 1 mL of TrypLE Select was added per T25 flask and incubated for 5 min at 37°C. The TrypLE was removed (cells should still be attached) and cells were flushed with 4 mL of complete DEF-CS™medium with growth factors 1-3, diluted 1:333, 1:1000, and 1:1000, respectively. The cell suspension was collected in a 15 mL conical tube and counted.

Cells were diluted to $2.1 * 10^5$ cells/mL corresponding to 42,000 cells per 200 µL (for 6 constructs + control in total need = $1.2 * 10^6$ cells in 5.6 mL). First, the required amount of OPTI-MEM was pipetted, then the DNA and FuGENE-HD were added (should not get in contact with the plastic of the tube), for exact transfection conditions see Table A.1 and A.2. The suspension was mixed well by vortexing and spun down quickly, and incubated for 15 min at RT. 800 µL of the cell suspension was added to each transfection mixture sample and resuspended gently. The coating was removed by aspiration and 200 µL of cell-transfection-suspension was transferred to each well for a seeding density of 42,000 cells/well, x3 replicates for each condition. The remaining non-transfected cells were plated as negative control (200 µL per well for 42,000 cells/well). Cells were incubated at 5% CO_2 at 37°C.

The media was changed once daily post-transfection. 2-3 days after transfection, cells were subcultured onto a new 24-well plate and treatment with 20 ng/mL diphtheria toxin (DT) proceeded with 3 replicates for each condition. During DT selection, the media was changed daily with complete DEF-CS™(with GF1 & 2) supplemented with 20 ng/mL DT. As DT selection generates lots of dead cells in the beginning, the plates were washed once with PBS−/− prior to the addition of fresh DT-containing medium. Once all control cells had died, the surviving cells in the transfected wells were allowed to recover, thus generating the finalized cell lines. The three generated ODIn-hiPSC lines (ODIn-U6-PCSK9, ODIn-U6-VEGFA2 and ODIn-U6-HEK4) were frozen in 1 ml/vial (x 3 biological replicates) of a specialized ready-to-use freezing medium for pluripotent stem cells at a concentration of $1 - 2 * 10^6$ cells/mL and stored at -80°C until further expansion and downstream differentiation applications.

Table A.1: Overview of constructs and plasmid dilutions for transfections

Plasmid	**Conc.** (ngµL^{-1})	**Plasmid** (µL)	**EB buffer**	**Vol.** (µL)	**New conc.**
EF1a-AzCas9*	1800	5.6	44.4	50	200
pBB282 (Xential sgRNA)	460	5.4	19.6	25	100
HDR donor U6-PCSK9	682.6	18.3	31.7	50	250
HDR donor U6-VEGFA**	833.2	15.0	35.0	50	250
HDR donor U6-HEK293**	743.7	16.8	33.2	50	250

*AzCas9, an AZ-proprietary Cas-enzyme, was used for the insertion of the guide cassettes. Had SpCas9 been used instead, editing at the guide target sites would have been induced, since the guides are continuously expressed. **The U6/iU6-VEGFA guides targets VEGFA site 2 (*VEGFA2*), likewise, the U6/iU6-HEK293 guides targets HEK293 site 4 (abbreviated *HEK4*)

A.3 ODIn-hiPSC culture

Having received the cryovials of ODIn-hiPSCs with clonally selected and inducible CRISPR-SpCas9 constructs, the first task was to thaw and culture the cells in suitable conditions for downstream differentiation applications.

Table A.2: Transfection set-up (all volumes in µL)

	250 ng DNA (1 rxn)	**1000 ng DNA** (for 4 rxn)
Opti-MEM	11.17	44.68
EF1a-AzCas9 (200 ng µL^{-1})	0.33	1.32
HBEGF-targeting gRNA (100 ng µL^{-1})	0.20	0.80
HDR donor (250 ng µL^{-1})	0.67	2.68
FuGENE HD	0.625	2.50

A.3.1 ODIn-hiPSC thawing & seeding into the Cellartis® DEF-CS™500 System

The ODIn cell cultures were initially thawed and seeded in preparation for the first, partial trilineage differentiation to endodermal-, mesodermal- and ectodermal cells. New ODIn-hiPSCs were then thawed and seeded after validation of differential potential (QC), for initiation of cell-type directed differentiation into hepatocytes, ventricular cardiomyocytes, and midbrain dopaminergic neurons. The kit used for the ODIn-hiPSC thawing, seeding, and initial cell expansion was the Cellartis®DEF-CS™500 Culture System (cat. nr. Y30010; nr. Y30012; nr. Y30017, Takara Bio) with the accompanying protocol (*Cellartis®DEF-CS™500 Culture System User Manual*), as summarized below. The Cellartis®DEF-CS™500 Culture System is an easy-to-use, rigorously validated, and complete culture system for defined and feeder-free ES and iPS cell expansion in a non-colony, 2D-monolayer format.

Three Corning®Costar®6-well flat bottom and tissue culture-treated plates (cat. nr.CLS3516, MERCK) were used for thawing and seeding of the ODIn-hiPSC, one plate (and well) per cell line. The frozen ODIn-hiPSC cryovials were collected from the −80°C freezer and transported to the biosafety cabinet fully submerged in dry ice. 1.5 mL Cellartis®DEF-CS 500 COAT-1 (cat. nr. Y30012, Takara Bio) per cell line was prepared, diluted 1:20 in Gibco™D-PBS (10x) +/+ (cat. nr. 14080055, ThermoFisher). The diluted COAT-1 was mixed well by pipetting up and down x3-4 times and finally added to one well per plate. The coating was spread out by applying quick side-to-side motions to the plates and tapping them against the bench for removing any remaining air bubbles. The plates were then incubated in a cell incubator at 37°C, 5% CO_2, and > 90% humidity for a minimum of 20 minutes for settling. 7 ml/cell line of complete Cellartis®DEF-CS™Medium for Thawing and Passaging was prepared by diluting Cellartis®DEF-CS™growth factors 1-3, 1:333, 1:1000 and 1:1000, respectively, with DEF-CS™Basal Media (cat. nr. Y30017, Takara Bio). The following steps for cell thawing and seeding with the Cellartis®DEF-CS™System for PSC culture proceeded one ODIn-hiPSC line at a time and each step was repeated subsequently. The frozen ODIn-hiPSC samples were partially submerged (not letting the opening of the lid pass the surface for risk of contamination) in a water bath at 37°C and stirred around until only a small piece of ice remained in the vial (1-3 min). Once thawed, 4 mL Cellartis®DEF-CS™Medium for Thawing and Passaging was mixed with the cell suspension (1 ml) using a 5 mL stripette. The diluted cell suspension was directly transferred to a

50 mL conical tube (5 mL liquid in total) and centrifuged for 5 min at 300xg in RT. The supernatant was aspirated, leaving enough medium to not disrupt the cell pellet. The bottom of the tube was gently flicked to loosen the cell pellet, which was subsequently resuspended in fresh DEF-CS™Medium for Thawing and Passaging (3 ml/sample). Excessive COAT-1 was aspirated from the culture ware directly before seeding the cells. The cell suspension was gently mixed by pipetting up and down 2-3 times with a 5 mL stripette and then directly transferred to the coated well of the culture ware. Cells were dispersed over the well by applying a circular ("8-shape") motion to the plate directly before incubation at 37°C, 5% CO_2, and $> 90\%$ humidity overnight.

A.3.2 Maintaining ODIn-hiPSC cultures in the Cellartis® DEF-CS™500 System

Full media changes proceeded daily on subsequent days after seeding (except on Saturday, when seeding on a Monday), as well as checking the cells in a microscope for ensuring healthy cell morphology. The old media was aspirated and fresh Cellartis®DEF-CS™medium for Maintenance (GF1 diluted 1:333 & GF2 diluted 1:1000 in DEF-CS™Basal Media), heated to 37°C in a water bath prior to use, was added with a pipette to the seeded well of the tilted culture ware, avoiding flushing the cells directly.

When the ODIn-hiPS cultures reached $> 75\%$ confluency ($1.5-3.0*10^6$ cells/mL), as indicated by the yellow shift in color of the phenol-containing DEF-CS™media and by visually checking the cell density in a standard microscope, the cells were passaged to new 6-well flat bottom and TC-treated plates (Corning). Passage occurred on day 3 or 4 after seeding or previous passage, following to the steps outlined in the *Cellartis®DEF-CS™System User Manual* (Takara Bio) but excluding the optional Step 5. The protocol proceeded one ODIn-hiPSC line at a time and each step was repeated subsequently. 1.5 ml/well COAT-1 was prepared by dilution in D-PBS +/+ (1:20) and added to each well of the culture ware (3x6-well plates, one for each ODIn-hiPSC line). The coated plates were incubated at 37°C, 5% CO_2, and $> 90\%$ humidity for at least 20 min before seeding. 3 ml/well seeding (with an additional 3 ml/cell line of DEF-CS™Medium for Thawing & Passaging for resuspension and seeding) was prepared through dilution of GF1-3 in DEF-CS™Basal Medium (1:333, 1:1000 & 1:1000).

The old medium was aspirated from the bottom of each well, carefully tilting the plate to not disturb the cells. Each well was washed once with 2 mL Gibco™PBS (1X, pH 7.4) without Mg^{2+} or Ca^{2+} (−/−) (cat. nr.10010023, ThermoFisher) and then 300 µL of Gibco™TrypLE™Select Enzyme (1X) w.o phenol red (cat. nr.12563011) was added to the bottom of the tilted vessel to gently disassociate the cells without directly flushing them with the trypsin-reagent. The TrypLE-treated cells were subsequently incubated at 37°C, 5% CO_2 and $> 90\%$ humidity for approximately 5 min (until the cells had visibly detached from the wells). Meanwhile, 10 µL Trypan blue was prepared in 0.5 mL Eppendorf tubes (1 per ODIn-hiPSC line) for cell counting with the Invitrogen™Countess™Automated Cell Counter. For the final passage before trilineage differentiation, the more accurate Vi-CELL™XR Cell

Viability Analyzer (Beckman-Coulter) was used for cell counting. Then, 1 mL of the resuspended cells, diluted 1:2 or 1:4 in fresh DEF-CS™Media for Thawing and Passaging was prepared.

After 5 min of incubation, the TrypLE™was carefully removed with a pipette (not diluted as suggested by the protocol), to avoid prolonged exposure to the dissociation reagent while the plate was tilted to prevent cells from being discarded with the media. The cells were resuspended in 3 ml/well-harvested of DEF-CS™Medium for thawing and passaging. To resuspend the cells, the bottom of each treated well was flushed with 3x1 mL of fresh media followed by pipetting up and down 3-4 times (only for the first mL of media to avoid too much shear force on the cells) in a circular and forceful motion. The cell suspension was then directly transferred to a 50 mL conical tube (one per cell line/condition) and the plate was discarded after ensuring that no cells were remaining. The resuspended cells were mixed with a pipette 2-3 times for disrupting any remaining cell aggregates. For cell counting with the Countess™, 10 µL of cell suspension was mixed with the prepared aliquot of Trypan blue. 10 µL of the well-mixed solution was finally added to a glass slide for counting. Total cell count was used when counting with the Countess™(low sample volume), contrarily, only the live cell count was used when counting with the Vi-CELL™. In both cases, the percentage of live cells should be as high as possible and the total cell count between $1.5 - 3.0 * 10^6$ cells/mL at confluency.

The remaining cell suspension was diluted in fresh DEF-CS™media for Thawing and Passaging, to obtain a new seeding density of $4 - 5x10^4$ cells/cm^2, the lower amount was for waiting 4 days in-between passages and the higher was for waiting 3 days. As many wells as possible were seeded but more importantly, the appropriate seeding density. Finally, excess COAT-1 was aspirated from the newly prepared culture ware and 3 ml/well of diluted cell suspension was seeded. The seeded wells were marked and the new passage number was noted, together with the cell line label and date. The cells were incubated overnight at 37°C, 5% CO_2 and $> 90\%$ humidity.

A.3.3 Maintaining ODIn-hiPSC cultures in the mTeSR™1 System

In preparation for the iPSC differentiations with the kits from STEMCELL™Technologies (trilineage assay, ventricular cardiomyocytes, and midbrain neurons), the ODIn-hiPSCs were transferred from the Cellartis®DEF-CS™500 System to the mTeSR™1 System (cat. nr. 85850, STEMCELL™Technologies) with Corning®Matrigel®Matrix Solution (cat. nr.CLS354234, MERCK). mTeSR™1 is a widely used feeder-free maintenance medium for human embryonic stem cells (hESCs) and human induced pluripotent stem cells hiPSC and is also more compatible than Cellartis DEF-CS™with the company's own differentiation kits. Preparation of complete mTeSR™1 proceeded as described in Section 4.1 of the *Technical Manual: Maintenance of Human Pluripotent Stem Cells in mTeSR1* (STEMCELL™Technologies). In summary, mTeSR™1 5X Supplement was thawed overnight at $2-8$°C and warmed to room temperature (RT, $15 - 25$°C) before use. To prepare 500 mL of complete mTeSR™1 medium, 100 mL of the 5X supplement was diluted in 400 mL

mTeSR™1 Basal Medium and mixed thoroughly. For the media change, the old DEF-CS™medium for maintenance was aspirated and replaced with 3 ml/well of complete mTeSR™1 warmed to RT.

Corning®Matrigel®Coating ("CMC"), was prepared according to the protocol provided in Section 4.2.2 of *Technical Manual: Maintenance of Human Pluripotent Stem Cells in mTeSR1.* The CMC was kept on ice for the entire duration of thawing and handling, to prevent gelling. For 24 mL CMC, one aliquot (270 µL) of CMC concentrate was thawed on ice. 24 mL of cold DMEM/F-12 (STEMCELL™Technologies) was dispensed into a 50 m conical tube together with the thawed aliquot of CMC, and mixed well. The diluted CMC was immediately dispensed in the desired culture ware, 125 $\mu L\,cm^{-2}$. The CMC was dispersed evenly across the surface by swirling and to get rid of air bubbles each plate was tapped against the LAF-bench. Finally, the culture-ware was sealed with Parafilm®to prevent evaporation of the CMC and stored in the fridge at $2-8°C$ overnight. 30 min before use, bring coated culture ware to room temperature ($15-25°C$).

A.4 hiPSC differentiation

A.4.1 Trilineage differentiation (pluripotency QC)

To assess the pluripotency of the ODIn-hiPSCs used in this study the STEMdiff™Trilineage Differentiation Kit (cat. nr.05230, STEMCELL™Technologies) was used. The trilineage kit includes specialized, complete media and monolayer-based protocols to perform parallel *in vitro* directed differentiation experiments for each germ layer, establishing trilineage differentiation potential within one week.

A.4.1.1 Media and preparations

One day prior to initiation of the trilineage differentiation protocol for assessment of differentiation capacity of the three ODIn-hiPSC lines, the DEF-CS™Medium for maintenance was switched to mTeSR™1 (STEMCELL™Technologies). By first switching media from DEF-CS™to mTeSR™1, but keeping the cells in the same COAT-1-treated culture ware, the cells were gradually adjusted to the new system, which reduces the risk of cytotoxic stress. Preparation of complete mTeSR™1 and Corning®Matrigel®Coating (CMC) proceeded as described in Section A.3.3. In summary, the old DEF-CS™medium for maintenance was aspirated and replaced with 3 ml/well of complete mTeSR™1 warmed to RT. Moreover, CMC for 10 wells x 9 Corning®Costar®24-well flat bottom and TC-treated plates (3 cell lines x 3 cell lineages each) was prepared. Finally, the culture ware was sealed with Parafilm®to prevent evaporation of the CMC and stored in the fridge at $2-8°C$ overnight together with the frozen STEMdiff™Trilineage Ectoderm-, Mesoderm-, and Endoderm Media, for thawing.

A.4.1.2 Part A - Plating

On Day 0 of the STEMdiff™trilineage differentiation assay, the ODIn-hiPSCs were passaged from the Cellartis®DEF-CS™500 Culture System to the Corning®Matrigel®-coatedculture ware. The instructions provided by the *STEMdiff™Trilineage Differentiation Kit User Manual from STEMCELL™technologies* were followed without significant deviations. Note that the protocol's recommendations for plating format (24-well plates) were adopted so the experimental details provided here are for passaging human ES/iPS cells from a 6-well plate format to set up the STEMdiff™trilineage differentiation assay in 24-well plates. See Table A.3 below for a summary of the seeding densities, media, and protocol duration for each germ-layer differentiation.

Table A.3: Summary of the set-up for the trilineage differentiation assay

Cell Lineage	**Seeding Density** (cells/cm^2)	**Cells per Well** (24-well format)	**Vol. SCPM (Day 0)** (ml/well)	**Vol. Diff. Media** (ml/well)	**Differentiation Media**	**Protocol Duration** (days)
Endoderm	2×10^5	4×10^5	0.5	1	STEMdiff™Trilineage Endoderm Media (Cat. nr. 05231, STEMCELL™Technologies)	5
Mesoderm	4×10^5	8×10^5	0.5	1	STEMdiff™Trilineage Mesoderm Media (Cat. nr. 05232, STEMCELL™Technologies)	5
Ectoderm	5×10^4	1×10^5	0.5	1	STEMdiff™Trilineage Ectoderm Media (Cat. nr. 05233, STEMCELL™Technologies)	7

The coated culture ware was brought to room temperature ($15 - 25$°C) inside a biosafety cabinet 30 minutes before use and the Parafilm®was removed. Sufficient volumes of mTeSR™1 (3 ml/cell line and condition), DMEM/F-12 (1 ml/well harvested), PBS $-/-$ (1 ml/well harvested) and 1 ml/well harvested of Gentle Cell Dissociation Reagent were aliquoted and warmed to RT. Single-Cell Plating Medium (SCPM) was prepared by diluting Rho-associated protein kinase (ROCK)-inhibitor, Y-27632 (cat. nr.72302, STEMCELL™Technologies), 1:1000 in mTeSR™1, reaching a final concentration of 10 µmol L^{-1}. Each well to be passaged was washed once with 1 mL of PBS$-/-$. 1 ml/well of GCDR was added to the cells after aspiration of the PBS, followed by incubation at 37°C for 10 minutes. Meanwhile, one 50 mL conical tube per cell line was prepared with 1 mL of DMEM/F-12 per well harvested. After incubation, cells were dislodged in the dissociation reagent by pipetting up and down 4-5 times, flushing the surface in a circular motion with a 1 mL pipette tip. Cell aggregates were broken up and resuspended, and the single-cell suspension

was immediately transferred to the prepared conical tube. Harvested wells were subsequently washed with 2x1 ml/well DMEM/F-12 to collect any remaining cells and transferred to the conical tube. Cells were centrifuged at 300 x g for 5 minutes in RT, the supernatant aspirated and the cells resuspended in 1 mL of the prepared SCPM.

The cells were prepared for counting with the Vi-Cell XR automated cell counter (Beckman-Coulter) by diluting 250 µL of cell suspension in 750 µL SCPM (4x dilution). The viable cell count was used for the final calculation of the dilution ratio for seeding the cells into the new plating format. Immediately prior to seeding, the excess CMC was carefully aspirated from the prepared culture ware. 0.5 ml/well of cells diluted in SCPM was added to each well at a seeding density indicated in Table A.3. In total, 3 wells per cell lineage were plated for the ODIn-U6-VEGFA2 and ODIn-U6-HEK4 cell lines and 6 wells per cell lineage for the ODIn-U6-PCSK9 cell line, due to a significantly higher cell count at passage in this culture.

A.4.1.3 Part B - Differentiation

One day after seeding the cells into the nine 24-well, CMC-plates, monolayer trilineage differentiation was initiated through a complete media change, switching from SCPM to one of the three complete STEMdiff™Trilineage Differentiation Media (endodermal, mesodermal or ectodermal). Three plates for each differentiation protocol (one for each ODIn-cell line). 1 ml/well seeded of thawed media was warmed to room temperature ($15-25°C$). The old SCPM was aspirated from the incubated cell cultures and 1 ml/well of the appropriate STEMdiff™Trilineage Differentiation Medium was added with a 1 mL pipette to each well (3 wells/cell lineage for ODIn-U6-VEGFA2 & ODIn-U6-HEK4, 6 wells/cell lineage for ODIn-U6-PCSK9). Cells were incubated at 37°C for 24 hours and in-between complete, daily media changes on subsequent days. The differentiation assay proceeded until day 5 for mesoderm and endoderm lineages, and day 7 for ectoderm lineages (all cell lines), when the cells were ready for assessment of trilineage differentiation potential.

A.4.1.4 Part C - Dox-titration assay

The dox dose-response evaluation was performed on the remaining trilineage differentiated ODIn-PCSK9 cells (3x3 wells, 12-well plate format, Corning®) at day 5 of differentiation for the endodermal and mesodermal cells, and at day 7 for the ectodermal cells. On the day of harvest, the old medium was aspirated and exchanged for 1 ml/well of dox-spiked STEMdiff™differentiation media, at four different concentrations (one biological replicate each); 0.1, 1 & 10 ng/mL for ectodermal cells and 1, 10 & 100 ng/mL for endodermal and mesodermal cells. After 48h of incubation at 37°C, cells were harvested and DNA simultaneously extracted with 100 µL/well QE (Lucigen™) and incubated at 37°C for 10 min. After incubation, cells were transferred to 1.5 mL Eppendorf tubes (one tube per treated condition) and vortexed for 15s. Each QE-DNA sample was transferred to PCR tubes and incubated in a thermocycler at 70°C for 10 minutes, followed by 98°C for 10 min. The reaction was finally cooled down to 4°C and stored in $-20°C$ overnight. Targeted deep sequencing by AmpSeq was then performed to analyze the editing efficiency

(indel/SNP) in the on-target loci (*PCSK9*, *VEGA2* or *HEK4*), for each respective dox dose. For further details on AmpSeq library preparation see Section A.5.

Dox-dilutions:

For preparing the dox-spiked media, a serial dilution of was performed. First, 1 µL of doxycycline stock solution (10 mg/mL) was added to 999ul of Dulbecco's Modified Eagle's Medium/Nutrient Ham's Mixture F-12 with 15 mM HEPES buffer (DMEM/F-12, cat. nr.36254, STEMCELL™Technologies), to reach a new concentration 10μ/mL (dilution 1). Then for 100 ng/mL, 10 µL of the first dilution was added to 990 µL of a lineage-specific differentiation media (endo/meso/ecto, cat. nr.05230, STEMCELL™Technologies). Subsequently, for diluting to 10 ng/mL, 1 µL of the first dilution was added to 999 µL of lineage-specific differentiation media. For 1 ng/mL, 1 µL of the first dilution was added to 9 µL of DMEM/F-12 in a 0.5 mL Eppendorf tube, to a concentration of 1 µg/mL (dilution 2). Then, 1 µL of this second dilution was added to 999 µL of lineage-specific differentiation media. Finally, for 0.1 ng/mL, 1 µL of the second dilution was added to 9 µL of lineage-specific differentiation media.

A.4.2 Cell-type directed differentiations

Table A.4 summarizes the different differentiation kits used in this study.

A.4.2.1 Generation of ventricular cardiomyocytes

Part A - Preparation of media

Sterile technique was used to prepare STEMdiff™Ventricular Cardiomyocyte Differentiation Media (Differentiation Basal Medium + Differentiation Supplement A, B, and C, respectively). Differentiation Supplements A, B, and C were thawed at room temperature ($15-25°$C) and mixed thoroughly by swirling the flasks. 10 mL of each differentiation supplement was added to 90 mL of Differentiation Basal Medium, respectively, and the three media stocks were mixed thoroughly. The resulting STEMdiff™Ventricular Cardiomyocyte Differentiation Media (A, B, and C) were stored at $2-8°$C for up to 2 weeks and warmed to room temperature before use.

STEMdiff™Cardiomyocyte Maintenance Medium (Maintenance Basal Medium + Maintenance Supplement) was prepared using a sterile technique. The maintenance Supplement was thawed at room temperature ($15-25°$C) and mixed thoroughly. 10 mL of Maintenance Supplement was added to 490 mL of Maintenance Basal Medium, and the mixture was mixed thoroughly. The resulting STEMdiff™Cardiomyocyte Maintenance Medium was stored at $2-8°$C for up to 4 weeks and warmed to room temperature before use.

Part B - Differentiation

Cardiomyocyte differentiation started with the three ODIn-hiPSC cultures with high expression of the pluripotency marker SSEA4, maintained on CMC 6-well plates, as described in Section 3.2. 3x12-well tissue culture plates coated with CMC were prepared a day in advance (Day -3) and kept in the fridge ($2-8°$C) wrapped in Parafilm®overnight. On Day -2 of the protocol, cells were dissociated with the GCDR protocol, see Section A.1.2 and the cell pellet gently resuspended in 1 mL of

Table A.4: Cell-specific differentiation kits & protocol time

Cell type	Supplier	Kit	Additional needs	Protocol time
CMs	STEMCELL™Tech.	STEMdiff™Ventricular Cardiomyocyte Differentiation Kit (cat. nr.05010, STEMCELL™Tech.)	needs Y-27632	15 days
Hepatocytes	Takara Bio	Cellartis®Definitive Endoderm Differentiation Kit (cat. nr. Y30030, Takara Bio Europe)	-	6 days
	Takara Bio	Cellartis®Hepatocyte Differentiation Kit (cat. nr. Y30050, Takara Bio Europe)	-	14 days
Neurons	STEMCELL™Tech.	STEMdiff™SMADi neuronal Induction Kit (cat. nr.05835, STEMCELL™Tech.)	-	19-22 days
	STEMCELL™Tech.	STEMdiff™Midbrain Neuron Differentiation Kit (cat. nr.100-0038, STEMCELL™Tech.)	Human Recombinant Shh (cat. nr.C24II, STEMCELL™Tech.)	6-9 days
	STEMCELL™Tech.	STEMdiff™Midbrain Neuron Maturation kit (cat. nr.100-0041, STEMCELL™Tech.)	-	14 days

mTeSR™1 supplemented with 10 µmol L^{-1} Y-27632. A cell count was performed using the Vi-CELL XR Cell Analyzer (Beckman Coulter). Excess CMC was aspirated from the pre-coated 12-well plates, and 1 ml/well seeded of mTeSR™1 supplemented with 10 µmol L^{-1} Y-27632 was added. The ODIn-hiPSCs were diluted to two different seeding densities per cell line, I. $3.5 * 10^5$ cells/well and II. $4.5 * 10^5$ cells/well, and seeded on a 12-well plate each (six plates in total). All six plates were then moved in several quick, short, back-and-forth and side-to-side motions to ensure uniform distribution of the cells. The plate was subsequently incubated at 37°C for 24 hours without disturbing the cells. On Day -1, the medium was removed and replaced with 1 mL of fresh mTeSR™1 (without Y-27632) and incubated at 37°C for 24 hours without disturbing the cells. Cells were assessed for confluency on Day 0 and must reach >95% confluency before starting the differentiation protocol.

On Day 0 of the CM differentiation protocol, Matrigel®was thawed on ice and 20 µL was added to 2 mL of STEMdiff™Ventricular Cardiomyocyte Differentiation Medium A. The medium was then added (2 ml/well) to each of the wells on the six 12-well plates and the plates were incubated at 37°C. From Day 2 to Day 14 of the protocol, a full-medium change was performed every 2 days as follows: the

medium was gently removed from the wells using a pipettor, without aspirating it, and replaced with 2 mL of medium per well (Media B on Day 2, Media C on Day 4 and 6, followed by Maintenance Media on Day 8, 14 & 16).

A.4.2.2 Generation of hepatocytes with the Takara Bio protocol

Part A - Preparation of media

The Cellartis iPS Cell to Hepatocyte Differentiation System (cat. nr. Y30055, Takara Bio Europe AB) is a complete system for the differentiation of human induced pluripotent stem (iPS) cells to hepatocytes via definitive endoderm (DE) differentiation. This kit is optimized for hiPSCs maintained and expanded in the Cellartis®DEF-CS™Culture System and includes the Cellartis Definitive Endoderm Differentiation Kit (cat. nr. Y30030, Takara Bio Europe) for differentiation to DE cells, see *Cellartis Definitive Endoderm Differentiation Kit with DEF-CS™Culture System User Manual*, and subsequently the Cellartis Hepatocyte Differentiation Kit (cat. nr. Y30050, Takara Bio Europe) for subsequent differentiation to hepatocytes, see *Cellartis Hepatocyte Differentiation Kit User Manual*. The Cellartis DE Differentiation Kit contains frozen complete media and frozen coating solution for the differentiation of hPSC to DE cells in a monolayer format. All media components and coatings were thawed and prepared according to the manufacturer's instructions.

Part B - Differentiation

On day 0, cell culture vessels were coated using the ready-to-use Definitive Endoderm coating solution. ODIn-hiPSCs were dissociated into a single cell suspension, re-suspended in warm DE diff Day 0 medium, and seeded with a density of $4 * 10^4$ cellscm^{-2} (in 0.2 mL mediumcm^{-2}) in 6-well TC-treated cell culture plates (Corning™costar™). On days 1, 2, 3, 4 and 6, medium changes were performed using warm DE diff Day 1, 2, 3, 4, and 6 medium, respectively. On day 7, the DE cells were enzymatically dissociated with TrypLE select and re-seeded in 96- and 24-well TC-treated cell culture plates, respectively (Corning™costar™). First, the cell culture vessels were coated. To this end, thawed Hepatocyte Coating (Cellartis Hepatocyte Diff Kit; Takara Bio Europe AB; Y30050) was added directly to cell culture vessels (0.15 mL$\,\mathrm{cm}^{-2}$), and incubated for $\geq$ 60 min at RT. Then, the DE cells were washed with PBS $-/-$, warm TrypLE Select was added (0.1 mL$\,\mathrm{cm}^{-2}$), and incubated for 3–5 min at 37°C. Then 10% FBS in PBS $-/-$ was added (0.1 mL$\,\mathrm{cm}^{-2}$) to achieve a 1:1 dilution of the cell suspension. Next, the cell suspension was centrifuged for 5 min at 300 g at RT, the supernatant was removed, and the cell pellet was re-suspended in warm Hepatocyte Thawing and Seeding medium (Cellartis Hepatocyte Diff Kit; Takara Bio Europe AB; Y30050). Then, the excess coating was removed and $1.3{*}10^5$ DE cellscm^{-2} were seeded in 0.5 mL mediumcm^{-2}. On days 9 and 11 (counted from the start of hiPSC differentiation), full medium changes were performed using Hepatocyte Progenitor Medium (Cellartis Hepatocyte Diff Kit). On days 14 and 16 media changes were performed using Hepatocyte Maturation Medium (Cellartis Hepatocyte Diff Kit) and only 90% of the media was removed so as not to disturb the gelatinous overlay. An IR-thermometer was used to ensure that the temperature of the maturation media was in the specified range of $16 - 18$°C. From day 18 onwards, media changes were performed every second or

third day using Hepatocyte Maintenance Medium (Cellartis Hepatocyte Diff Kit), only removing 90% of the media and carefully so as not to disturb the overlay.

A.4.2.3 Generation of neurons

The STEMdiff™Midbrain Neuron Differentiation kit (Cat nr., STEMCELL™Tech.) is a two-part protocol for differentiating iPSCs into midbrain neurons. It allows the user to choose between generating 3D neuronal organoids via the formation of embryoid bodies, or a monolayer neuronal differentiation protocol. In this study, the latter was performed as it is a faster protocol.

Part A - Preparation of media

STEMdiff™SMADi neuronal Induction Kit consists of a defined, serum-free medium and supplement for efficient neuronal induction of human embryonic stem (ES) cells and induced pluripotent stem (hiPS) cells. The kit combines STEMdiff™neuronal Induction Medium (cat. nr.05835, STEMCELL™Technologies) with STEMdiff™SMADi neuronal Induction Supplement, which directs differentiation by blocking TGF-β/BMP-dependent SMAD signaling, resulting in efficient neuronal induction of even hard-to-differentiate cell lines. The neuronal induction media was thawed and prepared with SMADi, according to the manufacturer's instructions. The STEMdiff™Midbrain Neuron Differentiation Kit (Catalog nr.100-0038) generates midbrain neuronal precursors from neuronal progenitor cells (NPCs) derived from human pluripotent stem cells (hPSCs) using STEMdiff™SMADi neuronal Induction Kit. The midbrain neuronal precursors are further matured into midbrain neurons using STEMdiff™Midbrain Neuron Maturation Kit (Catalog nr.100-0041).

Part B - Differentiation

After two weeks of iPSC expansion in the DEF-CS™500 System, a second passage was performed and the three ODIn-hiPSC lines switched to mTeSR™1 (STEMCELL™Technologies), to adjust the cells to the STEMdiff™culture conditions. neuronal induction was initiated by reseeding the cells using the GCDR dissociation protocol, and resuspending cells at a density of $2 * 10^5$ cellscm^{-2} in warm (37°C) STEMdiff neuronal induction medium supplemented with 0.5 mL SMADi supplement (neuronal induction kit) and 10 µmol L^{-1} Y-27632 (2 ml/well). The hiPSC were seeded on PLO/Laminin coated 6-well plates (Corning®Costar®) that had been warmed to RT before use. Daily media changes were performed with STEMdiff neuronal induction medium supplemented with SMADi for 7 days when all cultures were approximately 90% confluent. Cells were passaged in the same neuronal induction culture system three times, at a density of $1 - 1.25 * 10^5$ cellscm^{-2}, with daily media changes (neuronal induction media + SMADi) in between. Y-27632 was only supplemented to the induction media upon passage. On day 21 (counting from starting neuronal induction of the hiPSCs), neuronal precursor cells (NPCs) were obtained. The NPCs were then either directly differentiated towards midbrain dopaminergic neurons with the STEMdiff Midbrain Neuron Differentiation and Maturation Kit (STEMCELL™Technologies) or maintained in STEMdiff™neuronal progenitor medium (STEMCELL™Technologies) for one week before initiating differentiation. On day 0 of neuron differentiation, the NPCs were passaged to freshly coated (PLO/Laminin) 6-well plates at a density of $1 - 1.25 * 10^5$ cellscm^{-2} in 2 ml/well of either neuronal induction media or neuronal progenitor me-

dia, depending on previous culture conditions. On day 1, a full media change with warm 2 ml/well of complete STEMdiff™Midbrain Neuron Differentiation Medium was performed. Complete differentiation media was prepared by supplementing STEMdiff™Midbrain Neuron Differentiation Basal Medium with thawed Midbrain Neuron Supplement and Human Recombinant Shh at a concentration of 200 ng/mL (STEMCELL™Technologies). Daily media changes with neuron differentiation media (2 ml/well) proceeded for one week, or until the cultures were 90% confluent. On day 8, the neuronal precursors were passaged to freshly coated (PLO/Laminin) 96-well plates at a density of $4-6*10^4$ cellscm^{-2} with the Accutase™dissociation protocol. Cells were resuspended in STEMdiff™Midbrain Neuron Maturation Medium (BrainPhys™Neuronal Medium + Maturation Supplement). The neurons were matured for a minimum of two weeks, with half media changes every two to three days, using complete neuron maturation media.

A.4.3 Cell characterization assays

A.4.3.1 Gene expression profiling with quantitative PCR (qPCR)

Comparative gene expression analysis with real-time quantitative PCR (qPCR) was performed on extracted mRNA from the trilineage differentiation assay and cell type-directed differentiation assays, to evaluate ODIn-hiPSC pluripotency, and finally to evaluate the cell type-directed differentiations of the mature ventricular cardiomyocytes, hepatocytes, and midbrain-dopaminergic neurons, respectively.

qPCR preparation

Three biological replicates per cell line and differentiation condition were harvested (one well eq. one biological replicate). Harvesting was performed with the RLT protocol, see Section A.1.2 for details. Total mRNA content was extracted using the Qiagen AllPrep®Mini- or Micro Kits (cat. nr.80204; nr.80284, Qiagen). mRNA was eluted with 30 µL nuclease-free water (cat. nr. AM9937, Invitrogen™, ThermoFisher), and the RNA-concentration measured with a NanoDrop™One/One^C Microvolume UV-Vis Spectrophotometer (cat. nr.ND-ONE-W, Thermo Scientific™). ODIn-hiPCS mRNA (without CRISPR guide-RNA) was provided by A. Madsen as a negative control for all qPCR assays. 20 ng total RNA per reaction was reverse-transcribed to cDNA with the High-Capacity cDNA Reverse Transcription Kit with RNase Inhibitor (cat. nr.4374966, Applied™Biosystems). If necessary, the mRNA was diluted in RNase-free water to a total volume of 13.2 µL per reaction. See the complete list of reagents per cDNA reaction below;

cDNA master mix (MM)	**vol. per reaction** [µL]
10x RT buffer	2
25x dNTP Mix (100mM)	0.8
10X Rrt Random Primers	2
RT	1
Rnase Inhibitor	1
RNA [µL]	13.2
Total volume [µL]:	20

The cDNA reactions were prepared on ice and the reverse-transcriptase was kept in a temperature-controlled cooling box to avoid unwanted reaction. After preparing the cDNA reactions, the following cDNA synthesis protocol was run on a thermocycler; 25°C for 10 min, then 37°C for 120 min, then 85°C for 5 min and finally hold at 4°C.

The cDNA was subsequently assayed with qPCR using cell lineage-specific TaqMan™MGB-binding Fluorescent Probes (Applied™Biosystems) on a QuantStudio 7 Flex qPCR machine (Applied Biosystems™). Three technical replicates per sample were assayed in a 384-well format, with 2-3 different lineage-specific probes, as well as two housekeeping genes (all cell types) and also a pluripotency marker-gene for the characterization assays with cell-type directed differentiated cells. In all runs, control ODIn-hiPSC mRNA was included (provided by A. Madsen) and assayed with the same probes as the differentiated cells.

The choice of probes was based on recommendations from each respective differentiation protocol if provided. When such recommendations were absent, the probes were chosen with reference to previously published literature. A detailed description of the customized TaqMan™probes used for qPCR in both the trilineage differentiation- and the cell-type directed differentiation assays are provided in Table A.5 below. In each qPCR reaction, 4 µL cDNA were assayed with 6 µL of TaqMan™Gene Expression Master Mix (Applied Biosystems™) prepared by diluting 5 µL MM with 0.5 µL of a lineage-specific, commercial TaqMan MGM-binding fluorescent probe (Applied Biosystems™) and 0.5 µL nuclease-free water.

The following qPCR protocol was run on the QuantStudio 7 Flex machine (Thermo Fisher Scientific™); Hold at 95,0 °C for 20 s, then run 40 cycles of the following program; 95,0°C for 1 s and 60,0°C for 20 s.

Data analysis

The output C_T-values were exported from the QuantStudio system's Design & Analysis software and transformed to relative (fold-change) gene expression values following standard Comparative C_T-analysis ($2^{-\Delta\Delta C_T}$ calculations). Marker-specific C_T-values were normalized against C_T-values for both housekeeping genes (PGK1 and GAPDH), respectively, and corresponding hiPSC control C_T-values and converted to $2^{-\Delta\Delta C_T}$-values.

A.4.4 Immunostaining and flow cytometry analysis

A.4.4.1 Antibodies and staining protocols

Table A.6 gives an overview of all antibodies used for the different cell characterization assays with immunostaining and flow cytometry analysis. Note that all antibodies used were conjugated with fluorescent reporters (fluorophores), either phycoerythrin (PE), Allophycocyanin (APC) or Fluorescein isothiocyanate (FITC).

For the immunostaining, the manufacturer's own protocols for internal/surface stainings (Miltenyi and Abcam) were used, specifically for each antibody. First, cells were harvested according to one of the aforementioned cell dissociation protocols, see Section A.1.2 and counted using the Vi-CELL XR Cell Analyzer (Beckman Coulter). $1 * 10^5$ cells were used per cell line and staining condition.

Miltenyi surface staining protocol

Table A.5: List of TaqMan™Probes used for quantitative gene expression analysis by qPCR

Sample	Probes	Assay	Reference
Endodermal cells	SOX17, CXCR4 & FOXA2	Hs00751752_s1, Hs00237052_m1 & Hs00232764_m1	STEMdiff™Trilineage Differentiation Kit (STEMCELL™Technologies)
Mesodermal cells	Brachyury (T), NCAM & CXCR4	Hs00610080_m1, Hs00941830_m1 & Hs00237052_m1	STEMdiff™Trilineage Differentiation Kit (STEMCELL™Technologies)
Ectodermal cells	Pax6 & Nestin	Hs01088114_m1 & Hs04187831_g1	STEMdiff™Trilineage Differentiation Kit (STEMCELL™Technologies)
ODIn-hiPSC/pluripotency control assays	POU5F1 & NANOG	Hs04260367_H & Hs02387400_1	[109]
Cardiomyocytes (ventricular)	MYH6, NKX2-5 & TNNT2	Hs01101425_m1, Hs00231763_m1 & Hs00165960_m1	[7, 37, 90]
Hepatocytes	HNF4A, A1AT (SERPINA1), ALB & FOXA2	Hs00230853_m1, Hs00165475_m1 & Hs00609411_m1	[14, 30] & *STEMdiff™Hepatocyte Kit* (STEMCELL™Technologies)
Neurons (mid-brain dopaminergic)	TUBB3, FOXA2 & TH	Hs00964962_g1, Hs00232764_m1 & Hs00165941_m1	[108] & *STEMdiff™Midbrain Dopaminergic Neuron Differentiation and Maturation Kits* (STEMCELL™Technologies)
Control assays	PGK1 & GADPH	Hs00943178_g1 & Hs02758991_g1	[9, 81]

A buffer solution containing phosphate-buffered saline (PBS −/−), pH 7.2, 0.5% bovine serum albumin (BSA), and 2 mM EDTA was prepared by diluting MACS® BSA Stock Solution (cat. nr. 130-091-376, Miltenyi Biotech) 1:20 with autoMACS® Rinsing Solution (cat. nr. 130-091-222, Miltenyi Biotech). The buffer and antibodies were kept cold (and Abs protected from light) on ice for the entire staining procedure. All steps were performed in the absence of artificial light to protect the fluorophores. Additionally, the incubation steps were performed in the dark. $1 * 10^5$ cells were centrifuged at 300g for 10 minutes and the supernatant aspirated completely. Cells were resuspended in 98 µL of buffer. The desired antibody (2 µL) was added and the mixture was incubated for 10 minutes under aluminium foil in the fridge ($2 - 8°C$). The cells were washed by adding 1-2 mL of buffer and centrifuging at 300g for 10 minutes, and the supernatant was aspirated completely. Finally, the cell pellet was resuspended in 100 µL buffer for flow cytometry analysis. For more details see *Cell surface flow cytometry staining protocol (PBS/EDTA/BSA 1:50)* (Miltenyi Biotech).

Miltenyi intracellular staining protocol

The same buffer solution as for the Miltenyi surface staining protocol was used for the Miltenyi intracellular staining procedure and all reagents were kept on ice for the entire staining procedure and all steps were performed in the same light conditions as described above. $1 * 10^5$ cells were resuspended in 250 µL of buffer and 250 µL of Inside Fix was added to the mixture. After mixing well, the cells were incubated for

Table A.6: List of conjugated antibodies used for immunostaining and flow cytometry analysis

Target	Intracellular (I) or extracellular/surface (S) protein	Antibody and fluorescent conjugate	Positive control cell line	Company	Product nr.	Staining protocol
(no specific target)	S	REA Control Antibody (I) human IgG-PE	N/A	Miltenyi Biotech	nr.130-118-347	Either staining protocols
Stage-specific embryonic antigen-4 (SSEA4)	S	human SSEA4-PE	ODIn-hiPSC (without sgRNA cassette)	Miltenyi Biotech	130-122-958	Miltenyi surface staining protocol
Asialoglycoprotein receptor 1 (ASGPR1)	S	ASGPR1-PE	Hepatocellular carcinoma cell line HepG2	Miltenyi Biotech	130-122-963	Abcam surface staining protocol
Albumin (ALB)	I	ALB-FITC	Hepatocellular carcinoma cell line HepG2	Abcam		Abcam intracellular staining protocol
(no specific target)	S	Goat polyclonal IgG-FITC	N/A	Abcam		Either staining protocols
Cardiac Troponin T (cTnT)	I	human cTnT-PE	Osteosarcoma cell line U2-OS	Miltenyi Biotech	130-120-545	Miltenyi intracellular staining protocol
(no specific target)	S	Mouse monoclonal IgG2a-APC	N/A	Abcam	ab91364	Either staining protocols
Beta III Tubulin (TUBB3)	I	TUBB3-APC	Neuroblastoma cell line SH-SY5Y	Abcam	ab224977	Abcam intracellular staining protocol

1 hour in the dark at room temperature. The mixture was then centrifuged at 300g for 5 minutes, and the supernatant aspirated. The cells were washed by adding 1 mL of buffer and centrifuging at 300g for 5 minutes, and the supernatant was carefully aspirated. The cells were then washed by adding 1 mL of Inside Perm (Inside Stain Kit) and centrifuging at 300g for 5 minutes. The supernatant was again carefully aspirated. The resuspended cells were resuspended in 98 µL of Inside Perm, and the desired antibody (2 µL) was added. The solution was mixed well and incubated for 1 h in the dark at room temperature. The cells were then washed by adding 1 mL of Inside Perm and centrifuging at 300g for 5 minutes, and the supernatant was carefully aspirated. Finally, the cell pellet was resuspended in a suitable amount of buffer for flow cytometry analysis and stored at $2 - 8°C$ in the dark until analysis. It is important to note that samples may be stored at $2 - 8°C$ in the dark for up to 24 hours, and they should be mixed well before flow cytometric acquisition. For more details see *Intracellular flow cytometry staining protocol (Inside Stain Kit 1:50)* (Miltenyi Biotech).

Abcam surface staining protocol

A buffer solution was prepared by diluting BSA to 0.5% (v/v) in Hanks' Balanced Salt Solution (HBSS). The buffer and all other reagents were kept on ice for the entire staining procedure and all steps were performed in the same light conditions as previously described. $1 * 10^5$ cells were centrifuged at 500g for 5 minutes, and the supernatant completely removed. Cells were resuspended in 98 µL of buffer, and 2 µL of the desired antibody was added. The solution was mixed well and incubated for 10 minutes in the dark in the refrigerator ($2 - 8°C$). The cells were washed by adding 1-2 mL of buffer and centrifuging at 300g for 5 minutes, and the supernatant was completely aspirated. The cell pellet was resuspended in 300-400 µL of HBSS for analysis. When staining clumpy cells (cardiomyocytes and hepatocytes), they

were passed through a cell strainer prior to flow cytometry acquisition.

Abcam intracellular staining protocol

$1 * 10^5$ cells were washed twice with 2 mL of HBSS with 0.5% (v/v) BSA. After each wash, the cells were centrifuged at 500g for 5 minutes, and then the buffer was removed from the pelleted cells. 500 µL of cold Flow Cytometry Fixation Buffer (cat. nr.FC004, Abcam) was added to the cells and the mixture was vortexed. The cells were incubated at room temperature for 10 minutes while being vortexed intermittently to maintain a single-cell suspension. The cells were centrifuged and the Fixation Buffer was removed. The cells were washed twice with HBSS as described in the first step. The cells were resuspended in 100 µL of Flow Cytometry Permeabilization Buffer/Wash Buffer I (cat. nr.FC005, Abcam). The desired antibody was added to the cell-containing buffer (5 µL TUBB3-Alexa488/IgG-Alexa488 or 2 µL cTnT-PE/IgG-PE), tubes were vortexed and then incubated for 30-60 minutes at room temperature under aluminum foil. After incubation, cells were washed twice with Flow Cytometry Permeabilization Buffer/Wash Buffer I (cat. nr.FC005, Abcam) as described in the first step. The cells were finally resuspended in 300-400 µL HBSS for flow cytometric analysis. Clumpy cells were passed through a cell strainer prior to flow cytometric acquisition.

A.4.4.2 Flow cytometry analysis

For the flow cytometry analysis, a FACSymphony™A1 Flow Cytometer (BD®) was used according to the manufacturer's instructions. The reason for choosing the FACSymphony™A1 Flow Cytometer (BD) for the flow cytometry analysis is that it was the best available machine, but nevertheless, reportedly a powerful tool for high-parameter flow cytometry that offers improved sensitivity, resolution, and workflow efficiency compared to others (BD). Below is a summary of the flow cytometric acquisition procedure and data analysis performed. For full setup and analysis details, see *FACSymphony™A1 Flow Cytometer User's Guide*.

Data acquisition

After staining the cells with fluorescently-conjugated antibodies, the FACSymphony™A1 Flow Cytometer (BD) and accompanying software (BD FACSDiva) were started and prepared for acquisition by priming the fluidics and running a quality control analysis with BD®CS&T beads, to prepare the optics and electronics. A new experiment was created in BD FACSDiva and the desired lasers and analysis parameters such as side-scatter (SSC) and forward-scatter (FSC) were selected. When ready for acquisition, the positive control samples were run, followed by isotype control (IgG), and finally the samples of interest (lineage-specific stains). As the cells pass through the instrument, they are illuminated by a laser, and the emitted fluorescence is detected and recorded by the machine.

Gating

The data collected by the flow cytometer was analyzed using the FACSymphony A1 software, the *BD FACSDiva*. The first step in the analysis was to set gates, which define the population of cells that will be analyzed in subsequent steps. Gating was based on forward and side scatter (which provide information about cell size and granularity, respectively), as well as on the fluorescence intensity (PE or AF488, respectively) of the stained cells.

Data analysis and interpretation

Once the gates were set, the fluorescence intensity of the stained cells was analyzed to provide information about the expression levels of the proteins targeted by the antibodies. Finally, the acquired data were interpreted in the context of the experimental question. In this case, identifying subpopulations of cells with distinct phenotypes based on their protein expression patterns. More specifically, the parameter of interest for the cell characterization assays was the percentage of cells in the total cell population that was positive for a lineage-specific target protein or pluripotency marker. This percentage was weighed against the corresponding percentage of positive cells in the positive control samples, where a high percentage (close to 100%) was expected. Also, it was of interest to assess the subpopulation positive for the isotype control (should be close to 0%).

A.4.5 Induction of CRISPR-SpCas9 genome editing in ODIn hiPSC-derived cell lineages

For the different hiPSC-derived cell lineages, different doxycycline (dox) doses and incubation times were applied. In the following sections, the dox-dilution preparation protocols for the different cell lines are presented. See a summary of the doxycycline induction regiments below:

A.4.5.1 Cardiomyocytes

Initial Dox-concentration: 10 mg/mL = 10,000 μg/mL
Final concentration: 100 ng/mL = 0.1 μg/mL
Number of wells induced: 8 (ODIn-V) + 8 (ODIn-P) + 12 (ODIn-H) = 28 wells $\times$ 2 mL CM Maintenance media $\rightarrow$ prepared 60 mL dox-spiked media

$$C1 * V1 = C2 * V2$$

Dilution 1: Diluted 1 µL of 10 mg/mL Dox into 999 µL CM Maintenance Media ($\times$1000) $\rightarrow C\_1 = 10$ mg/mL, dilute $\times$100.
Dilution 2: $V_1 = \frac{C\_2V_2}{C\_1} = \frac{0.1 \times 60}{10}$ mL = 0.6 mL = 600μl into 59.4 mL CM Maintenance media.

Media change:

Warmed the 60 mL aliquot of dox-spiked media and a 60 mL aliquot of unspiked CM Maintenance Media to RT. Split the wells per cell type into two groups, mixed wells so that there were dox-treated and controls from each seeding density condition. Performed a regular media-change with CM Maintenance Media on the wells marked as controls C and a media change with Dox-spiked Maintenance Media on the wells marked Dox. Cells were incubated at 37°C for 48h before removing the dox-spiked and unspiked media and performing a regular media change with fresh, untreated CM maturation media. Cells were subsequently maintained in unspiked maintenance media for another week, with media changes every 2 days, before harvest with the STEMdiff™Cardiomyocyte Dissociation Protocol (STEMCELL™Technologies) and preparing for subsequent cell characterization (immunostaining/qPCR) and on/off-target gene editing assays.

A.4.5.2 Hepatocytes

The same step-wise dilution procedure as applied for the cardiomyocytes, was performed to prepare dox-spiked aliquots for inducing SpCas9 expression and subsequent genome editing in the hepatocytes (both batches). Yet, here the final dose desired was 250 ng/mL dox, diluted in either of the two hepatocyte maintenance media (Cellartis®, Takara Bio or STEMdiff™, STEMCELL Tech.), depending on the hepatocyte batch. Also, one unspiked aliquot for the control samples was prepared. Then a regular (but full instead of 90% as specified by the Takara Bio protocol) media change was performed (0.5 $mL\,cm^{-2}$). Caution was taken to not disturb the gelatinous overlay in the Takara Bio-hepatocyte culture. Three to four wells per ODIn-cell line were induced with doxycycline and three to four replicates were kept as negative controls (without dox). Cells were incubated at 37°C for 48h before removing the dox-spiked and unspiked media and performing a regular media change with fresh, untreated HP maintenance media. Cells were subsequently maintained in unspiked maintenance media for another week, with media changes every 2 days, before harvest with the TrypLE Select 1x protocol (Gibco™) and preparing for subsequent cell characterization (immunostaining/qPCR) and on/off-target gene editing assays.

A.4.5.3 Neurons

Performed the same stepwise dilution procedure as for the cardiomyocytes and hepatocytes, however, prepared one aliquot of 1000 ng/mL dox-spiked neuron maturation media and one unspiked aliquot for the control samples (STEMCELL™Technologies). At least four wells per ODIn-cell line were induced with doxycycline and at least four replicates were kept as negative controls (without dox). Cells were incubated at 37°C for 72h before removing the dox-spiked and unspiked media and performing a regular media change with fresh, untreated NR maturation media. Cells were subsequently maintained in unspiked maturation media for another two weeks, with media changes every 2-3 days, before harvest with the Accutase®protocol (Innovative Cell Technologies, Inc) and preparing for subsequent cell characterization (immunostaining/qPCR) and on/off-target gene editing assays. After a week, a genotyping assay was however performed on the neurons, as a mix-up of the cell lines was suspected.

A.4.6 Genotyping ODIn neurons

To confirm whether the two ODIn-hiPSC-derived neuron cell lines had been mixed up, somewhere during differentiation, or in fact, were mixed cultures, genotyping was performed. Three different assays were tested for this purpose; PCR, TIDE and AmpSeq.

A.4.6.1 PCR

Initially, PCR analysis with both *PCSK9*-specific- and VEGFA-site 2-specific primers was performed on three biological replicates per cell line of dox-induced ODIn neurons.

Primers:

Forward primer: seq-U6-fw (5′−GAGGGCCTATTTCCCATGATTCC−′3) Reverse primer: antisense oligo from VEGFA2/PCSK9 gRNAs

Set-up:

Prepared PCR reactions according to Table A.7. Two master mixes with cell line-specific primers were prepared, excluding the DNA. In total, 12 reactions were prepared, 6 with PCSK9-rv primers and 6 with VEGFA-rv primers so that each cell line could be assayed with both primers; S1) $ODIn - PCSK9_{PCSK9} - rep1$, S2) $ODIn - PCSK9_{PCSK9} - rep2$, S3) $ODIn - PCSK9_{PCSK9} - rep3$, S3) $ODIn - VEGFA_{PCSK9} - rep1$, S4) $ODIn - VEGFA_{PCSK9} - rep2$, S5) $ODIn - VEGFA_{PCSK9} - rep3$, S6) $ODIn - VEGFA_{VEGFA} - rep1$, S7) $ODIn - VEGFA_{VEGFA} - rep2$, S8) $ODIn - VEGFA_{VEGFA} - rep2$, S9) $ODIn - VEGFA_{VEGFA} - rep3$, S10) $ODIn - PCSK9_{VEGFA} - rep1$, S11) $ODIn - PCSK9_{VEGFA} - rep2$ and S12) $ODIn - PCSK9_{VEGFA} - rep3$.

Table A.7: PCR set-up for genotyping ODIn neurons

Per reaction	**Volume (µL)**
2x DreamTaq mix	5
Forward primer (10 µmol L^{-1})	1
Reverse primer (10 µmol L^{-1})	1
DNA	1
Water	2
Total	**10**

The following PCR program was run, see Table A.8. Afterward, 10 µL of each PCR product was loaded on a 1.5% agarose gel, and stained with Invitrogen™SYBR™Safe DNA Gel Stain (cat. nr.S33102, ThermoFischer) to confirm the correct product size. The expected fragment size is 274 bp.

Table A.8: PCR program for genotyping ODIn neurons

Step	**Temperature (°C)**	**Time**	
Initial denaturation	95	3 min	
Denaturation	95	30 sec	30 cycles
Annealing	60	30 sec	
Extension	72	1 min	
Final extension	72	5 min	
Hold	4	Indefinetly	

A.4.6.2 TIDE analysis

Procedure:

As a next attempt for genotyping the ODIn neurons, a Sanger sequencing-based TIDE analysis was performed. The library preparation (PCR) was performed by the

author, the Sanger sequencing was performed by Genewiz, and the TIDE analysis was performed by the author on the AZ internal sequencing analysis platform. See Table A.9 for all library preparation details. Note that the primer concentration (fw/rv is 10 µmol L^{-1}).

Table A.9: TIDE-analysis PCR

Library preparation for TIDE (PCR)											
DNA samples	**PCSK9 (TD primer nr. 33+34)**					**DNA samples**	**VEGFA2 (TD primer nr. 37+38)**				
3 dox-treated P	6.5 µL	H2O	98 °C	10 sec		3 dox-treated P	9.5 µL	H2O	95 °C	3 min	
3 dox-treated V	10 µL	Phusion 2x MM	98 °C	1 sec		3 dox-treated V	12.5 µL	Kapa 2x HIFI MM	98 °C	20 sec	
1 control HEK4	1 µL	Fw primer	65 °C	5 sec		1 control HEK4	0.75 µL	Fw primer	69 °C	15 sec	
	1 µL	Rv primer	72 °C	15 sec	30 cycles		0.75 µL	Rv primer	72 °C	30 sec	35 cycles
	1.5 µL	DNA	72 °C	1 min			1.5 µL	DNA	72 °C	2 min	
	20 µL	**Total volume**	4 °C	hold			25 µL	**Total volume**	4 °C	hold	

A.4.6.3 Targeted deep sequencing

Finally, the same samples were submitted to targeted deep sequencing analysis as a last definitive attempt to genotype the neuronal specimens. See the following Section A.5 for further details on the library preparation protocol.

A.5 Targeted deep sequencing (AmpSeq)

Verification of gene editing, as a result of Cas9-induced double-stranded breaks and erroneous DNA repair, was performed with the short-read targeted next-generation sequencing method, amplicon sequencing (AmpSeq).

A.5.1 AmpSeq library preparation

The specific protocol used for AmpSeq-prep was developed in-house by the AZ NGS team, who also performs the actual sequencing. Fundamentally, the extracted DNA amplicons (*VEGFA2*, *HEK4*, or *PCSK9*) were purified and amplified in two rounds of PCR and bead cleanup, as library preparation of the chosen amplicon sequence prior to NGS with Illumina NextSeq500. This strategy uses sequence-specific primers with linked adapters and sample-specific barcodes for enabling multiplexing the AmpSeq to use a pool of several samples (biological replicates and different cell lineages). This streamlines the protocol, reduces technical variation and facilitates the downstream data analysis of the results.

A.5.1.1 DNA amplification - PCR 1

In the 1st PCR, the *PCSK9* site was amplified using the Phusion Hot Start II High-Fidelity PCR Master Mix (cat. nr. F565S, ThermoScientific™), while the sites targeted by the promiscuous guides, *VEGFA2* and *HEK4*, were amplified using

KAPA HiFi Hotstart Ready Mix (cat. nr. 7958927001, Roche). The PCR reaction mixture included 1.5 µL of genomic DNA (0.1-100 ng) as input, along with site-specific and barcoded primers attached to sequencing adapters:

5′-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-[amplicon specific forward primer]-3′, and

5′-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-[amplicon specific reverse primer]-3′

The sequencing adapters were used in a subsequent PCR reaction, to anneal indexing primers. A list of all primers used is provided in *Appendix B*. A typical master mix and PCR cycling numbers are shown below. Products were verified on a fragment analyzer. If the FA indicated primer dimer formation, the PCR was repeated with halved primer concentrations. If primer dimers were still present, the primer concentration was further reduced.

PCSK9-primers					**VEGFA2- & HEK4-primers**				
PCR mix per sample		**PCR program**			**PCR mix per sample**		**PCR program**		
6.5 µL	Nuclease-free water	98 °C	10 sec		9.5 µL	H2O	95 °C	3 min	
10 µL	Phusion 2x mastermix	98 °C	1 sec	30 cycles	12.5 µL	Kapa 2x HIFI	98 °C	20 sec	30 cycles
1 µL	Forward primer [10 µmol L^{-1}]	65 °C	5 sec		0.75 µL	Fw primer [10 µmol L^{-1}]	65 °C	15 sec	
1 µL	Reverse primer [10 µmol L^{-1}]	72 °C	15 sec		0.75 µL	Rv primer [10 µmol L^{-1}]	72 °C	15 sec	
1.5 µL	Qiagen Miniprep DNA	72 °C	1 min		1.5 µL	DNA [max. 100 ng μL^{-1}]	72 °C	2 min	
20 µL	**Total volume**	4 °C	hold		25 µL	**Total volume**	4 °C	hold	

A.5.1.2 Dilution, pooling & PCR cleanup

Uniquely barcoded samples were diluted to the same concentration (in ng μL^{-1}) and pooled prior to being bead-washed and quantified by fragment analysis. In total, 30 combinations of forward- and reverse barcodes were kept, 6 fw and 5 rv. As the barcodes were the same for all three guides, the different cell lines could be pooled together as well, which allowed for a total of 30 samples from all conditions to be pooled into one sample for sequencing. Each sample (maximum per cell type: 3 cell lines x 3 replicates x 3 cell lineages each = 27 samples in tot) was thus labeled with a unique set of fw and rv barcodes. The barcoded PCR products were pooled based on the least concentrated sample, where 10 µL of this was used as a reference for calculating dilution factors for the rest of the samples. PCR products were diluted in Qiagen Buffer EB (cat. nr. 19086) and pooled together. 30 µL of the pooled sample was mixed with 20 µL EB. The pooled PCR products were then purified and size-selected using an equal volume (50 µL) of Agencourt AmPure XP magnetic beads (cat. nr. A63880, Beckman Coulter).

To initiate the purification process, the Agencourt AmPure XP magnetic beads at room temperature were vortexed for 20 seconds and then poured into a reservoir. Next, 50 µL of these beads were added to each PCR sample at a ratio of 1:1 (beads to sample). To ensure proper binding of the DNA fragments to the beads, the mixture was gently pipetted and mixed at least 10 times. The samples were then incubated

at room temperature for 10 minutes. After the incubation, the PCR strip was placed on a magnetic plate and left for 5 minutes or until the solution became clear. Once the solution clarified, the supernatant was carefully removed and discarded, leaving a small amount (5 µL) to avoid disturbing the beads. While the strip remained on the magnetic plate, 200 µL of 80% ethanol was added to wash the beads. The strip was incubated for 30 seconds, allowing the ethanol to come into contact with the beads, and then all the ethanol was removed while the beads were still attracted to the magnet. This washing step was repeated once.

After the final wash, the PCR strip was taken off the magnet, and the beads were left to air dry. It was important to ensure that all the ethanol evaporated, making the beads appear dull rather than shiny. Care was taken not to let them dry excessively and crack, as this could lead to lower DNA yield. To elute the purified PCR products, 26 µL of elution buffer was added to the beads. The mixture was pipetted and mixed 10 times, and then incubated at room temperature for 5 minutes. Following the incubation, the PCR strip was placed back on the magnetic plate and left for 2 minutes or until the solution became clear. After the magnetic separation, 25 µL of the eluate, divided into two transfers of 12.5 µL each, was carefully transferred to a new tube, ensuring that no beads were carried over.

Finally, to verify the quality of the purified PCR products, a 2 µL aliquot of the sample was run on a Fragment Analyzer using the DNF-915 kit from Advanced Analytical Technologies. This analysis confirmed the correct size of the product, its specificity, and the absence of primer dimers. Using the concentration from smear analysis with the Fragment Analyzer, purified samples were diluted to 0.067 ng μL^{-1} in Buffer EB (Qiagen), reaching a final volume of 100 µL.

A.5.1.3 Indexing - PCR 2

A second PCR was run on the pools to index the samples. This was performed by members of the NGS team. Here, a set of sequencing primers (cat. nr. FC-131-1002, Illumina Nextera XT index kit), annealing to the adapters added in the first PCR, was used to index each pool to enable further pooling. Each sample has one N7xx primer and one S5xx primer that is unique for that sample so that each read can be identified after sequencing. All pools were amplified with Kapa HiFi Polymerase (cat nr. KK2602, Roche), using the below PCR master mix and cycling numbers.

PCR MasterMix KAPA HiFi:	
incl nuclease-free water + wt controls	per rxn [µL]
KAPA HiFi * 2x	12.5
NGS Primer mix	5
DNA	7.5
Total reaction volume	25
*KAPA HiFi Hotstart Ready Mix	

PCR products were cleaned up using Agencourt AmPure XP beads, same ratio, 1:1, as in the previous clean-up (=50 µL beads). Purified products were verified on Fragment Analyzer (DNF-915 kit) to ensure size, concentration, and purity. The average fragment size and concentration of pools were used to pool and dilute

PCR MasterMix KAPA HiFi:	
incl nuclease-free water + wt controls	per rxn [µL]
KAPA HiFi * 2x	12.5
NGS Primer mix	5
DNA	7.5
Total reaction volume	25
*KAPA HiFi Hotstart Ready Mix	

samples to a final molarity of 20 nM in a single library to be submitted for NGS. Samples were diluted in buffer EB with 0.1% tween 20. The tween helps the very low amount of sample to not attach to the tube walls.

A.5.1.4 Illumina Next-Generation Sequencing

The indexed- and pooled libraries were subjected to paired-end sequencing using Illumina®NextSeq 500 mid-output with a 150 bp read length. The sequencing output was analyzed using the internal AZ NGS analysis software by M. Firth and then further analyzed by the author in another internal NGS analysis software. Briefly, the reads were merged and mapped to the amplicon reference sequence. Variants (excluding the single nucleotide variations) were called with a minimum allele frequency of 0.1% and a minimum base quality of 25 Phred 33 scores. The resulting variants were analyzed using RIMA. Variants not overlapping the cutting window (±8 base pairs of the cut site) were excluded. Finally, the editing efficiencies were measured as the percentage of modified reads in mapped reads.

A.5.2 rhAmpSeq off-target editing analysis

For the off-target analysis, a similar approach as for the on-target editing validation was used, but using the rhAmpSeq assay for multiplexed targeted amplicon sequencing.

A.5.2.1 rhAmpSeq library preparation

This protocol provides instructions for rhAmpSeq library preparation for the three customized rhAmpSeq panels (PCSK9, VEGFA2 and HEK4). The output of this protocol is NGS-ready libraries from genomic DNA using rhAmpSeq-targeted sequencing reagents. The samples included in the rhAmpSeq assay (cell line, cell lineage and dox-treatment/control) are summarised in Table A.10 below.

Custom rhAmpSeq panels (IDT) containing target-specific forward (fw) and reverse (rv) primers for PCR 1 of the rhAmpSeq library preparation, were ordered from IDT. For the full list of rhAmpSeq primer sequences, see *Appendix B*, Section B.3. A 10X rhAmp Primer Pool was established by diluting the guide-specific and customized rhAmp primer pools to 50 µmol L^{-1} in IDT buffer pH 7.5. The primers were supplied in a plate format (dried) and were centrifuged before-, as well as after resuspension. Resuspended primers were also thoroughly vortexed. After resuspension, six rhAmp primer pool stock solutions had been generated $PCSK9_{fw}$,

Table A.10: Sample layout for the rhAmpseq off-target analysis assay

	1	2	3	4	**5**	6	7	8	9	10	11	12
A	PC1_cm	PC1_nr	hiPSC_PC1	HC1_cm	hiPSC_HC1	V1C_cm	V1C_nr	hiPSC_VC1	x	x	x	x
B	PC2_cm	PC2_nr	hiPSC_PC2	HC2_cm	hiPSC_HC2	V2C_cm	V2C_nr	hiPSC_VC2	x	x	x	x
C	PC3_cm	PC3_nr	hiPSC_PC3	HC3_cm	hiPSC_HC3	V3C_cm	V3C_nr	hiPSC_VC3	x	x	x	x
D	PC4_cm	PC4_nr	hiPSC_PC4	HC4_cm	hiPSC_HC4	V4C_cm	V4C_nr	hiPSC_VD1	x	x	x	x
E	PD1_cm	PD1_nr	hiPSC_PD1	HD1_cm	hiPSC_HD1	V1D_cm	V1D_nr	hiPSC_VD2	x	x	x	x
F	PD2_cm	PD2_nr	hiPSC_PD2	HD2_cm	hiPSC_HD2	V2D_cm	V2D_nr	hiPSC_VD3	x	x	x	x
G	PD3_cm	PD3_nr	hiPSC_PD3	HD3_cm	hiPSC_HD3	V3D_cm	V3D_nr	hiPSC_VD4	x	x	x	x
H	PD4_cm	PD4_nr	hiPSC_PD4	HD4_cm	hiPSC_HD4	V4D_cm	V4D_nr	x	x	x	x	x

$PCSK9_{rv}$, $VEGFA2_{fw}$, $VEGFA2_{rv}$, $HEK4_{fw}$ and $HEK4_{rv}$ (one fw and one rv per rhampSeq panel). The rhAmpSeq Library Kit included (besides customized rhAmp primers) two 4X PCR master mixes (Library Mix 1 and Library Mix 2) specifically formulated for use with rhAmpSeq panels, dilution buffers (IDTE pH 7.5 and IDTE pH 8.0) and dried-down (6 nmol) indexing primers for PCR 2 (i5 and i7). The library mixes contain the required enzymes and buffer components necessary for library construction. rhAmpSeq Index Primers allow for the incorporation of unique sample indexes to the rhAmpSeq primers, as well as the addition of P5 and P7 indexing sequences. This further allows for the multiplexing of samples (pooling) during sequencing to increase throughput. rhAmpSeq Index Primers were resuspended to 100 µmol L^{-1} by adding 60 50 µL of IDTE, pH 8.0. Resuspensions were Vortexed and then centrifuged. No dilution or further preparation was needed for the master mixes or other reagents, and all rhAmpSeq library prep. reagents were stored at –20°C and thawed to RT before use.

The library prep was rerun twice for optimization of the DNA amplification (increasing DNA output for pooling and sequencing). Each sample of genomic DNA, see Table A.10, was diluted to 0.91 ng/µL to 9.1 ng/µL using IDTE, pH 8.0. For rhAmp PCR 1, 5 µL 4X rhAmpSeq Library Mix 1, 2 µL 10X rhAmp PCR Panel—Forward Pool, 2 µL 10X rhAmp PCR Panel—Reverse Pool and 11 µL of diluted gDNA was pipetted to each reaction in a 96-well plate (10–100 ng total DNA input). PCR 1 was performed on a thermal cycler with a lid temperature of 105°C and using the following protocol: 95°C for 10 minutes, 10 cycles (initial prep) or 15 cycles (re-prep) of 95°C for 15 s followed by 61°C for 4 minutes. Thereafter 99.5°C for 15 minutes and finally 4°C indefinite hold.

PCR 1 products were purified through bead-wash (Agencourt AMPure XP beads, Beckman Coulter), according to the same protocol as used for AmpSeq library prep but with 30 µL beads per reaction. PCR 1 products were eluted in 15 µL of IDTE, pH 8.0 per reaction. See Section A.5.1.2. For the second Indexing PCR, 5 µL 4X rhAmpSeq Library Mix 2, 2 µL Indexing PCR Primer i5 (5 µmol L^{-1}), 2 µL Indexing PCR Primer i7 5 µL and 11 µL of each amplified sample elution into a new 96-well plate. See Table A.11 for an overview of the primer set-up. PCR 2 was performed on

a thermal cycler with a lid temperature of 105°C and using the following protocol: 95°C for 3 minutes, 18 cycles (initial prep) or 20 cycles (re-prep) of 95°C for 15 s followed by 60°C for 30 s and then 72 °C for 30 s. Thereafter 72°C for 1 minute and finally 4°C indefinite hold. PCR 2 products were purified through bead-wash, according to the same protocol as used previously, but with 20 µL beads per reaction. PCR 2 products were eluted in 22 µL of IDTE, pH 8.0 per reaction. Targeted deep sequencing was performed by J. Lindgren in the NGS team of the Translational Genomics Department on the Illumina NextSeq500 platform (Illumina®), according to the same instructions as provided in Section A.5.1.4. Initial data analysis was performed by M.Firth in the Quantitative Biology Department, on the internal AZ NGS analysis platform. Subsequent NGS analysis and assessment of cell-type effects on CRISPR-Cas9 off-target editing was performed by J. Stevrell.

Table A.11: Indexing PCR primer set-up

Sample	**i5 forward**	**i7 reverse**
PC1_cm	i5-1	i7-1
PC2_cm	i5-1	i7-2
PC3_cm	i5-1	i7-3
PC4_cm	i5-1	i7-4
PD1_cm	i5-1	i7-5
PD2_cm	i5-1	i7-6
PD3_cm	i5-1	i7-7
PD4_cm	i5-1	i7-8
PC1_nr	i5-2	i7-1
PC2_nr	i5-2	i7-2
PC3_nr	i5-2	i7-3
PC4_nr	i5-2	i7-4
PD1_nr	i5-2	i7-5
PD2_nr	i5-2	i7-6
PD3_nr	i5-2	i7-7
PD4_nr	i5-2	i7-8
hiPSC_PC1	i5-3	i7-1
hiPSC_PC2	i5-3	i7-2
hiPSC_PC3	i5-3	i7-3
hiPSC_PC4	i5-3	i7-4
hiPSC_PD1	i5-3	i7-5
hiPSC_PD2	i5-3	i7-6
hiPSC_PD3	i5-3	i7-7
hiPSC_PD4	i5-3	i7-8
HC1_cm	i5-4	i7-1
HC2_cm	i5-4	i7-2
HC3_cm	i5-4	i7-3
HC4_cm	i5-4	i7-4
HD1_cm	i5-4	i7-5
HD2_cm	i5-4	i7-6

HD3_cm	i5-4	i7-7
HD4_cm	i5-4	i7-8
hiPSC_HC1	i5-5	i7-1
hiPSC_HC2	i5-5	i7-2
hiPSC_HC3	i5-5	i7-3
hiPSC_HC4	i5-5	i7-4
hiPSC_HD1	i5-5	i7-5
hiPSC_HD2	i5-5	i7-6
hiPSC_HD3	i5-5	i7-7
hiPSC_HD4	i5-5	i7-8
V1C_cm	i5-6	i7-1
V2C_cm	i5-6	i7-2
V3C_cm	i5-6	i7-3
V4C_cm	i5-6	i7-4
V1D_cm	i5-6	i7-5
V2D_cm	i5-6	i7-6
V3D_cm	i5-6	i7-7
V4D_cm	i5-6	i7-8
V1C_nr	i5-7	i7-1
V2C_nr	i5-7	i7-2
V3C_nr	i5-7	i7-3
V4C_nr	i5-7	i7-4
V1D_nr	i5-7	i7-5
V2D_nr	i5-7	i7-6
V3D_nr	i5-7	i7-7
V4D_nr	i5-7	i7-8
hiPSC_VC1	i5-8	i7-1
hiPSC_VC2	i5-8	i7-2
hiPSC_VC3	i5-8	i7-3
hiPSC_VC4	i5-8	i7-4
hiPSC_VD1	i5-8	i7-5
hiPSC_VD2	i5-8	i7-6
hiPSC_VD3	i5-8	i7-7
hiPSC_VD4	i5-8	i7-8

20 µL per reaction of purified rhAmp products were quantified on a Fragment Analyzer (DNF-915 kit) to ensure size, concentration, and purity. The average fragment size and concentration of pools were used to pool and dilute samples to a final molarity of 20 µmol L^{-1} in a single library to be submitted for NGS. Samples were diluted in buffer EB with 0.1% tween 20. For an overview of the final samples submitted for NGS, see Tables A.12, A.13 and A.14.

Table A.12: PCSK9 rhAmpSeq panel

Sample	**Finished Library Fragment Size (bp)**	**Finished Library Concentration (ng µL^{-1})**	**Amt of Library (µL)**	**Amt of EB + Tween (µL)**	**Total EB + Tween (µL)**	**Desired Molarity (nM)**	**Total Volume per sample (µL)**
PC1_cm	318	18.6953	2.25	7.75	132.6	20	10
PC2_cm	321	9.9105	4.28	5.72		20	10
PC3_cm	322	11.4165	3.72	6.28		20	10
PC4_cm	323	8.9702	4.75	5.25		20	10
PD1_cm	321	8.3515	5.07	4.93		20	10
PD2_cm	317	12.5892	3.32	6.68		20	10
PD3_cm	320	6.4055	6.59	3.41		20	10
PD4_cm	314	22.1356	1.87	8.13		20	10
PC1_nr	316	18.5008	2.25	7.75		20	10
PC2_nr	319	15.9904	2.63	7.37		20	10
PC3_nr	318	22.1459	1.90	8.10		20	10
PC4_nr	320	21.6393	1.95	8.05		20	10
PD1_nr	319	19.2431	2.19	7.81		20	10
PD2_nr	318	14.4756	2.90	7.10		20	10
PD3_nr	316	10.1826	4.10	5.90		20	10
PD4_nr	317	14.5494	2.88	7.12		20	10
hiPSC_PC1	318	8.1009	5.18	4.82		20	10
hiPSC_PC2	324	9.8462	4.34	5.66		20	10
hiPSC_PC3	322	9.6382	4.41	5.59		20	10
hiPSC_PC4	334	4.0472	10.89	-0.89		20	10
hiPSC_PD1	332	3.3899	12.93	-2.93		20	10
hiPSC_PD2	330	3.9941	10.91	-0.91		20	10

Continued on next page

Table A.12 – continued from previous page

Sample	**Finished Library Fragment Size (bp)**	**Finished Library Concentration** ($\mathbf{ng\,\mu L^{-1}}$)	**Amt of Library** (µL)	**Amt of EB + Tween** (µL)	**Total EB + Tween** (µL)	**Desired Molarity (nM)**	**Total Volume per sample** (µL)
hiPSC_PD3	320	8.3934	5.03	4.97		20	10
hiPSC_PD4	307	37.822	1.07	8.93		20	10

Table A.13: VEGFA2 rhAmpSeq panel

Sample	**Finished Library Fragment Size (bp)**	**Finished Library Concentration** ($ng\,\mu L^{-1}$)	**Amt of Library** (µL)	**Amt of EB + Tween** (µL)	**Total EB + Tween** (µL)	**Desired Molarity (nM)**	**Total Volume per sample** (µL)
V1C_cm	320	26.9217	1.57	8.43	134.5	20	10
V2C_cm	324	22.0872	1.94	8.06		20	10
V3C_cm	324	19.4623	2.20	7.80		20	10
V4C_cm	324	17.7433	2.41	7.59		20	10
V1D_cm	322	14.1252	3.01	6.99		20	10
V2D_cm	321	19.2621	2.20	7.80		20	10
V3D_cm	320	22.227	1.90	8.10		20	10
V4D_cm	319	23.9352	1.76	8.24		20	10
V2C_nr	323	35.5114	1.20	8.80		20	10
V3C_nr	320	25.0067	1.69	8.31		20	10
V4C_nr	319	24.6831	1.71	8.29		20	10
V1D_nr	312	28.6307	1.44	8.56		20	10
V2D_nr	301	5.9772	6.65	3.35		20	10
V3D_nr	319	21.6466	1.95	8.05		20	10
V4D_nr	312	34.0682	1.21	8.79		20	10
hiPSC_VC1	318	20.8721	2.01	7.99		20	10
hiPSC_VC2	326	18.9693	2.27	7.73		20	10
hiPSC_VC3	320	18.6182	2.27	7.73		20	10
hiPSC_VD1	326	3.2529	13.23	-3.23		20	10
hiPSC_VD3	313	3.2032	12.90	-2.90		20	10

Table A.14: HEK4 rhAmpSeq panel

Sample	**Finished Library Fragment Size (bp)**	**Finished Library Concentration** ($\mathbf{ng\,\mu L^{-1}}$)	**Amt of Library** (µL)	**Amt of EB + Tween** (µL)	**Total EB + Tween** (µL)	**Desired Molarity (nM)**	**Total Volume per sample** (µL)
HC1_cm	324	13.7267	3.12	6.88	80.3	20	10
HC2_cm	326	11.5032	3.74	6.26		20	10
HC3_cm	326	10.8221	3.98	6.02		20	10
HC4_cm	329	14.0841	3.08	6.92		20	10
HD1_cm	325	10.3475	4.15	5.85		20	10
HD4_cm	324	11.9436	3.58	6.42		20	10
hiPSC_HC1	325	18.0938	2.37	7.63		20	10
hiPSC_HC2	328	14.6149	2.96	7.04		20	10
hiPSC_HC3	324	10.0392	4.26	5.74		20	10
hiPSC_HC4	326	12.6084	3.41	6.59		20	10
hiPSC_HD1	321	7.7459	5.47	4.53		20	10
hiPSC_HD2	319	6.1358	6.86	3.14		20	10
hiPSC_HD3	310	14.7855	2.77	7.23		20	10

B

DNA sequences

In this Appendix, a complete list of DNA sequences used for the targeted deep sequencing assays (AmpSeq/rhAmpSeq) are presented. All sequences are in the $5' - 3'$ direction.

B.1 DNA sequences ODIn-hiPSCs

Table B.1: DNA sequences for the three genomic on-target sites *PCSK9*, *VEGFA2*, and *HEK4* in ODIn-hiPSCs and downstream cell types

Target name	Guide sequence	PAM	Genomic reference sequence
PCSK9	TCATCCGCCCGGTACCGTGG	NGG	AGTTGCCCCATGTCGACTACATCGAGGAGGACTCCTCTGTCTTTGCCCAGAGCATCCCGTGGAACCTGGAGCGGATTACCCCTCCACGGTACCGGG CGGATGAATACCAGCCCCCCGGTAAGACCCCCATCTGTGCCCTGCCCCACCCCATCTGAGCTGAATCCATTTGCTCTGCCCTGGCCTGGCCTC CCTGCTGGTGGTTTCCACTTCTCGG
HEK4	GGCACTGCGGCTGGAGGTGG	NGG	TGGAAGGAAGGGAGGAAGGGCGAGGCAGAGGGTCCAAAGCAGGATGACAGGCAGGGGCACCGCGGCGCCCCGGTGGCACTGCGGCTGGAGGTGG GGGTTAAAGCGGAGACTCTGGTGCTGTGTGACTACAGTGGGGGCCCTGCCCTCTCTGAGCCCCCGCCTCCAGGCCTGTGTGTGTGTCTCCGTT CGGTTGAAAGG
VEGFA2	GACCCCCTCCACCCCGCCTC	NGG	GGACAGACAGACAGACACCGCCCCCAGCCCCAGCTACCACCTCCTCCCCGGCCGGCGGCGGACAGTGGACGCGGCGGCGAGCCGCGGGCAGGG GCCGGAGCCCGCGCCCGGAGGCGGGTGGAGGGGGTCGGGGCTCGCGGCGTCGCACTGAAACTTTTCGTCCAACTTCTGGGCTGTTCTCGCTTC GGAGGAGCCGTGGTCCGCGCGGGGGAAGCCGAGCCGAGCGGAGCCGCGAGAAGTGCTAGCTCGGGC

B.2 Amplicon sequencing primers

All primers used for targeted deep sequencing (on-target editing validation) by Amplicon-seq on the Illumina NextSeq500 platform contained a 5'-sequencing adapter linked to a unique 4-nucleotide long barcode, compatible with Nextera XT Index Kit v2 (Illumina):

5′-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-
$-[barcode]-$ [amplicon specific forward primer]-3′, and
5′-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-
$-[barcode]-$ [amplicon specific reverse primer]-3′

Table B.2: PCSK9 primers for targeted deep-sequencing (AmpSeq)

Primer name	Abbreviation	AmpSeq adapter	Barcode	Primer sequence	Full sequence
PCSK9 fw-1	21a	5'-TCGTCGGCAGCGTCAG ATGTGTATAAGAGACAG-3'	AAGC	5'-AGTTGCCCCA TGTCGACTAC-3'	5'-TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGAAGCAGTTGCCCCATGTCGACTAC-3'
PCSK9 fw-2	21b	5'-TCGTCGGCAGCGTCAG ATGTGTATAAGAGACAG-3'	CGTT	5'-AGTTGCCCCA TGTCGACTAC-3'	5'-TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGCGTTAGTTGCCCCATGTCGACTAC-3'
PCSK9 fw-3	21c	5'-TCGTCGGCAGCGTCAG ATGTGTATAAGAGACAG-3'	GTCA	5'-AGTTGCCCCA TGTCGACTAC-3'	5'-TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGGTCAAGTTGCCCCATGTCGACTAC-3'
PCSK9 fw-4	21d	5'-TCGTCGGCAGCGTCAG ATGTGTATAAGAGACAG-3'	ACAG	5'-AGTTGCCCCA TGTCGACTAC-3'	5'-TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGACAGAGTTGCCCCATGTCGACTAC-3'
PCSK9 fw-5	21e	5'-TCGTCGGCAGCGTCAGA TGTGTATAAGAGACAG-3'	CAGT	5'-AGTTGCCCCA TGTCGACTAC-3'	5'-TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGCAGTAGTTGCCCCATGTCGACTAC-3'
PCSK9 fw-6	21f	5'-TCGTCGGCAGCGTCAGA TGTGTATAAGAGACAG	TGTC	5'-AGTTGCCCCA TGTCGACTAC-3'	5'-TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGTGTCAGTTGCCCCATGTCGACTAC-3'
PCSK9 rv-1	22a	5'-GTCTCGTGGGCTCGGAG ATGTGTATAAGAGACAG-3'	AGCA	5'-CCGAGAAGTG GAAACCACCA-3'	5'-GTCTCGTGGGCTCGGAGATGTGTATAAGA GACAGAGCACCGAGAAGTGGAAACCACCA-3'
PCSK9 rv-2	22b	5'-GTCTCGTGGGCTCGGAG ATGTGTATAAGAGACAG-3'	CCTT	5'-CCGAGAAGTG GAAACCACCA-3'	5'-GTCTCGTGGGCTCGGAGATGTGTATAAGA GACAGCCTTCCGAGAAGTGGAAACCACCA-3'
PCSK9 rv-3	22c	5'-GTCTCGTGGGCTCGGAG ATGTGTATAAGAGACAG-3'	GAAC	5'-CCGAGAAGTG GAAACCACCA-3'	5'-GTCTCGTGGGCTCGGAGATGTGTATAAGA GACAGGAACCCGAGAAGTGGAAACCACCA-3'
PCSK9 rv-4	22d	5'-GTCTCGTGGGCTCGGAG ATGTGTATAAGAGACAG-3'	TTGG	5'-CCGAGAAGTG GAAACCACCA	5'-GTCTCGTGGGCTCGGAGATGTGTATAAGA GACAGTTGGCCGAGAAGTGGAAACCACCA-3'
PCSK9 rv-5	22e	5'-GTCTCGTGGGCTCGGAG ATGTGTATAAGAGACAG-3'	ACAG	5'-CCGAGAAGTG GAAACCACCA-3'	5'-GTCTCGTGGGCTCGGAGATGTGTATAAGA GACAGACAGCCGAGAAGTGGAAACCACCA-3'
PCSK9 rv-6	22f	5'-GTCTCGTGGGCTCGGAG ATGTGTATAAGAGACAG-3'	CAGT	5'-CCGAGAAGTG GAAACCACCA-3'	5'-GTCTCGTGGGCTCGGAGATGTGTATAAGA GACAGCAGTCCGAGAAGTGGAAACCACCA-3'

Table B.3: VEGFA2 primers for targeted deep-sequencing (AmpSeq)

Primer name	Abbreviation	AmpSeq adapter	Barcode	Primer sequence	Full sequence
VEGFA2 fw-1	15a	5'-TCGTCGGCAGCGTCAG ATGTGTATAAGAGACAG-3'	AAGC	5'-GGACAGACAG ACAGACACCG-3'	5'-TCGTCGGCAGCGTCAGATGTGTATAAGAG ACAGAAGCGGACAGACAGACAGACACCG-3'
VEGFA2 fw-2	15b	5'-TCGTCGGCAGCGTCAG ATGTGTATAAGAGACAG-3'	CGTT	5'-GGACAGACAG ACAGACACCG-3'	5'-TCGTCGGCAGCGTCAGATGTGTATAAGAG ACAGCGTTGGACAGACAGACAGACACCG-3'
VEGFA2 fw-3	15c	5'-TCGTCGGCAGCGTCAG ATGTGTATAAGAGACAG-3'	GTCA	5'-GGACAGACAG ACAGACACCG-3'	5'-TCGTCGGCAGCGTCAGATGTGTATAAGAG ACAGGTCAGGACAGACAGACAGACACCG-3'
VEGFA2 fw-4	15d	5'-TCGTCGGCAGCGTCAG ATGTGTATAAGAGACAG-3'	ACAG	5'-GGACAGACAG ACAGACACCG	5'-TCGTCGGCAGCGTCAGATGTGTATAAGAG ACAGACAGGGACAGACAGACAGACACCG-3'
VEGFA2 fw-5	15e	5'-TCGTCGGCAGCGTCAG ATGTGTATAAGAGACAG-3'	CAGT	5'-GGACAGACAG ACAGACACCG-3'	5'-TCGTCGGCAGCGTCAGATGTGTATAAGAG ACAGCAGTGGACAGACAGACAGACACCG-3'
VEGFA2 fw-6	15f	5'-TCGTCGGCAGCGTCAG ATGTGTATAAGAGACAG-3'	TGTC	5'-GGACAGACAG ACAGACACCG-3'	5'-TCGTCGGCAGCGTCAGATGTGTATAAGAG ACAGTGTCGGACAGACAGACAGACACCG-3'
VEGFA2 rv-1	16a	5'-GTCTCGTGGGCTCGGAG ATGTGTATAAGAGACAG-3'	AGCA	5'-GCCCGAGCTA GCACTTCTC-3'	5'-GTCTCGTGGGCTCGGAGATGTGTATAAGA GACAGAGCAGCCCGAGCTAGCACTTCTC-3'
VEGFA2 rv-2	16b	5'-GTCTCGTGGGCTCGGAG ATGTGTATAAGAGACAG-3'	CCTT	5'-GCCCGAGCTA GCACTTCTC	5'-GTCTCGTGGGCTCGGAGATGTGTATAAGA GACAGCCTTGCCCGAGCTAGCACTTCTC-3'
VEGFA2 rv-3	16c	5'-GTCTCGTGGGCTCGGAG ATGTGTATAAGAGACAG-3'	GAAC	5'-GCCCGAGCTA GCACTTCTC	5'-GTCTCGTGGGCTCGGAGATGTGTATAAGA GACAGGAACGCCCGAGCTAGCACTTCTC-3'
VEGFA2 rv-4	16d	5'-GTCTCGTGGGCTCGGAG ATGTGTATAAGAGACAG-3'	TTGG	5'-GCCCGAGCTA GCACTTCTC	5'-GTCTCGTGGGCTCGGAGATGTGTATAAGA GACAGTTGGGCCCGAGCTAGCACTTCTC-3'
VEGFA2 rv-5	16e	5'-GTCTCGTGGGCTCGGAG ATGTGTATAAGAGACAG-3'	ACAG	5'-GCCCGAGCTA GCACTTCTC-3'	5'-GTCTCGTGGGCTCGGAGATGTGTATAAGA GACAGACAGGCCCGAGCTAGCACTTCTC-3'
VEGFA2 rv-6	16f	5'-GTCTCGTGGGCTCGGAG ATGTGTATAAGAGACAG-3'	CAGT	5'-GCCCGAGCTA GCACTTCTC-3'	5'-GTCTCGTGGGCTCGGAGATGTGTATAAGA GACAGCAGTGCCCGAGCTAGCACTTCTC-3'

Table B.4: HEK4 primers for targeted deep-sequencing (AmpSeq)

Primer name	Abbreviation	AmpSeq adapter	Barcode	Primer sequence	Full sequence
HEK4 fw-1	17a	5'-TCGTCGGCAGCGTCAG ATGTGTATAAGAGACAG-3'	AAGC	5'-TGGAAGGAAG GGAGGAAGGG-3'	5'-TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGAAGCTGGAAGGAAGGGAGGAAGGG-3'
HEK4 fw-2	17b	5'-TCGTCGGCAGCGTCAG ATGTGTATAAGAGACAG-3'	CGTT	5'-TGGAAGGAAG GGAGGAAGGG-3'	5'-TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGCGTTTGGAAGGAAGGGAGGAAGGG-3'
HEK4 fw-3	17c	5'-TCGTCGGCAGCGTCAG ATGTGTATAAGAGACAG-3'	GTCA	5'-TGGAAGGAAG GGAGGAAGGG-3'	5'-TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGGTCATGGAAGGAAGGGAGGAAGGG-3'
HEK4 fw-4	17d	5'-TCGTCGGCAGCGTCAG ATGTGTATAAGAGACAG-3'	ACAG	5'-TGGAAGGAAG GGAGGAAGGG-3'	5'-TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGACAGTGGAAGGAAGGGAGGAAGGG-3'
HEK4 fw-5	17e	5'-TCGTCGGCAGCGTCAG ATGTGTATAAGAGACAG-3'	CAGT	5'-TGGAAGGAAG GGAGGAAGGG-3'	5'-TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGCAGTTGGAAGGAAGGGAGGAAGGG-3'
HEK4 fw-6	17f	5'-TCGTCGGCAGCGTCAG ATGTGTATAAGAGACAG-3'	TGTC	5'-TGGAAGGAAG GGAGGAAGGG-3'	5'-TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGTGTCTGGAAGGAAGGGAGGAAGGG-3'
HEK4 rv-1	18a	5'-GTCTCGTGGGCTCGGAG ATGTGTATAAGAGACAG-3'	AGCA	5'-CCTTTCAACC CGAACGGAGA-3'	5'-GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGAGCACCTTTCAACCCGAACGGAGA-3'
HEK4 rv-2	18b	5'-GTCTCGTGGGCTCGGAG ATGTGTATAAGAGACAG-3'	CCTT	5'-CCTTTCAACC CGAACGGAGA-3'	5'-GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGCCTTCCTTTCAACCCGAACGGAGA-3'
HEK4 rv-3	18c	5'-GTCTCGTGGGCTCGGAG ATGTGTATAAGAGACAG-3'	GAAC	5'-CCTTTCAACC CGAACGGAGA-3'	5'-GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGGAACCCTTTCAACCCGAACGGAGA-3'
HEK4 rv-4	18d	5'-GTCTCGTGGGCTCGGAG ATGTGTATAAGAGACAG-3'	TTGG	5'-CCTTTCAACC CGAACGGAGA-3'	5'-GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGTTGGCCTTTCAACCCGAACGGAGA-3'
HEK4 rv-5	18e	5'-GTCTCGTGGGCTCGGAG ATGTGTATAAGAGACAG-3'	ACAG	5'-CCTTTCAACC CGAACGGAGA-3'	5'-GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGACAGCCTTTCAACCCGAACGGAGA-3'
HEK4 rv-6	18f	5'-GTCTCGTGGGCTCGGAG ATGTGTATAAGAGACAG-3'	CAGT	5'-CCTTTCAACC CGAACGGAGA	5'-GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGCAGTCCTTTCAACCCGAACGGAGA-3'

B.3 rhampSeq panels

Three different rhAmpSeq primer pools were designed by the author and A. Madsen, and manufactured by IDT™. One primer pool with 83-143 primers was customized per guide-RNA (PCSK9, VEGFA2, and HEK4), targeting predicted genomic off-target (OT) sites identified by *in vitro* and *in silico* analysis. See Table B.5, **??** and B.7 respectively.

Table B.5: PCSK9 rhAmpSeq primer pool

Target	Chr. location	Guide sequence
PCSK9	chr1:55046465-55046700	TCATCCGCCCGGTACCGTGG
OT01	chr6:150226400-150226624	TCATCCTCCCAGTACAATGG
OT02	chr1:230855696-230855878	ATATCCACCCAGGACCGTGG
OT03	chr4:72852647-72852821	TCACACACCGGGTCCTGTGG
OT04	chr3:54918462-54918632	TCA-CAGCCCAGTACCATGG
OT05	chr11:36134791-36134969	TCATCTGCCTGGTACCATGA
OT06	chr11:133501252-133501437	CTAGCCGCACTGTACCGTGG
OT07	chr19:28155681-28155868	TGTTCCTCCCAGTACCGTGG
OT08	chr19:1271323-1271483	CGATCCCGTGGGTACCGTGG
OT09	chrX:118496183-118496400	CCGGCCGCCCGGCACGGCGG
OT10	chr2:204329464-204329656	TCCTCCTGCCAGTCCAGTGG
OT11	chr22:46406815-46407042	CCATCAGCTGTGTGCCGTGG
OT12	chr2:10387232-10387384	GGATGCCCACGGTACCGTGG
OT13	chr10:11910474-11910636	CCATCCTTCCGGTACAGTGG
OT14	chr2:9461998-9462155	TCCTAGGCCCGGCACGGTGG
OT15	chr19:41464013-41464204	CCATACTCCCCATATCGTGG
OT16	chr1:58784834-58784995	TCATCCTCCGGG-ACCGTGG
OT17	chr19:10625850-10626023	ATCTCCACCTGGTTCCGTGG
OT18	chr17:50817277-50817501	TCACCCTCCCCTGACCGAGG
OT19	chr5:179117387-179117545	ACATCAGCCTGGTACAGTGG
OT20	chr5:2132383-2132550	TCCTCCGTCCATTGCCTTGG
OT21	chr6:138796049-138796229	TATGAGGCCCGGTACAGTGG
OT22	chr22:46433497-46433698	CCCTCCCCACGGCACCATGG
OT23	chr11:72308003-72308177	CTATCCACCTGGCACCATGG
OT24	chr5:26655315-26655479	TCTACCACCCTATACTGTGG
OT25	chr10:124089433-124089595	GCCTCAGCCCAGTACAGTGG
OT26	chr6:39212522-39212709	AGCTCCACCCGGTCCCGTGG
OT27	chr2:219570605-219570834	TCCTTCTCCCAGTACAGTGT
OT28	chr17:80048523-80048724	CCAGCAGCCCAGCACCGAGG
OT29	chr2:42302578-42302799	AAATCAGCCAGGTGCGGTGG
OT30	chrX:24505190-24505362	CGATCCCACCTGTAACGTGG
OT31	chr1:47625972-47626132	TCAACAGTCTTGTACCTTGG
OT32	chr19:42066874-42067077	TCTTCTCCCCAGTAGGGTGG
OT34	chr12:42292496-42292664	ACATTAGCCAGGTATGGTGG

Table B.5: PCSK9 rhAmpSeq primer pool

Target	Chr. location	Guide sequence
OT35	chr3:150635936-150636125	CTATGTGCTAGGTACCGTGG
OT36	chr10:132452745-132452906	AGATCCACCCAGCACCGTGC
OT37	chr11:44896265-44896446	TCATCCCCACAGTACAATGG
OT38	chr19:56393417-56393586	CGAATCCACCGGTACCGTGG
OT39	chr22:41946784-41946969	GCCAGCGCGCGCTACCGTGG
OT40	chr20:652931-653093	TCATCCGGCCGCTGCCGTCC
OT41	chr4:61430412-61430614	TCTTACTCTCTGTACAGTGG
OT42	chr6:42315835-42316040	GCATCTTCCAGGTACCGTGA
OT43	chr11:30129434-30129613	TGATAAGCCCTGTACAGTTG
OT45	chr17:13771626-13771800	TAATCCTTCCAGGACCATGG
OT46	chr9:41012765-41012970	CCACCCACCCGATACCGTTC
OT48	chr2:27372285-27372455	AAATCCTCCCCGTACTGTGG
OT49	chr6:164042212-164042379	TAAGCCAACCGTTACCGTAG
OT50	chr14:93756845-93757005	GAACCCCCTGGGTACCGTGG
OT51	chr16:56630464-56630634	TCATGCCCCTTGTACAGTGG
OT52	chr1:221634978-221635203	TTATGTGCCAGGTACTGTGC
OT53	chr15:63113906-63114096	CCTTGCGCCTGATACAGTGG
OT54	chr20:482460-482690	CCTGCAGCCCTGTACCGTGG
OT56	chr14:62859067-62859261	TAATCAGCCCGCTACC-TGG
OT57	chrX:6729641-6729805	TGATAAGCCTGGTACTGTGG
OT58	chr1:22387961-22388131	TCACATGCCCAGAACCGTGG
OT59	chr1:1284238-1284431	CCCGCCTCGAGGTACCGTGG
OT60	chr1:3600756-3600939	CCATCCCCCCGGCACCAGGG
OT61	chr1:67527267-67527438	TTTTCTGCCAGGTACAGTGC
OT62	chr11:1081823-1081989	CCCTCCACCCGATACCGGGG
OT63	chr5:15330701-15330886	TAATCCGCCCCTTATCCTTG
OT66	chr16:78273375-78273560	TTCTCTGCCTGGTACCGTGG
OT67	chr8:70439250-70439431	TCACTGGCCAGGTACTGTGG
OT68	chr1:119311638-119311803	TCATCAGCCTGGGTCCCTGA
OT69	chr14:76920813-76921044	ATGACCCCCAGGTACCGTGG
OT71	chr14:103848429-103848613	CCATTCGCCCGG-ACCGCGG
OT72	chr7:31775967-31776125	TCACCCCCACAGTACAGGGG
OT74	chr22:48703937-48704168	CCATCCACCCCGTACCATGG
OT75	chr10:124487492-124487651	CAATCAGCCCGGCACGGTGG
OT76	chr14:97001006-97001198	TCAGCAGGGAGGTACAGTGG
OT78	chrX:40214110-40214285	TCATCAGCTGGGTACAGTGG
OT79	chr8:133004697-133004859	ACATCTGCCTGGTACCCTGG
OT80	chr10:44861773-44861993	TCATCCCCACTTTACAGAGG
OT81	chr2:101475086-101475278	TCATGCAGCCGTTACTGTGG
OT82	chr2:97572977-97573204	TCACCCACCGGGGACTGTTG
OT83	chr5:107557873-107558060	TCCTCCCACCTATACCGTGG
OT84	chr12:1261468-1261658	TGAGCAGCACAGTACAGTGG

Table B.5: PCSK9 rhAmpSeq primer pool

Target	Chr. location	Guide sequence
OT86	chr15:65816058-65816266	AACTCAGCCAAGTACCGTGG
OT87	chr5:31291565-31291807	TCAGCAGCCTGGCACAGTGC
OT88	chr17:68427088-68427249	CCACCCACCTGGCACCGTGG
OT89	chr10:76977732-76977910	TCATCCACCCAGTACTTTGG
OT90	chr11:1923683-1923919	TCATCCTCCGTGTGCCATGG
OT91	chr7:2121209-2121399	GCACCTGCCAGGCACCGTGG
OT92	chr19:6468320-6468481	TCATCCTCAGGGCTCAGTGG
OT93	chr12:131403892-131404073	TCATCAGACCCCTACCAGGG

Table B.6: VEGFA2 rhAmpSeq primer pool

Target	Chr. location	Guide sequence
VEGFA	chr6:43770706-43770893	GACCCCCTCCACCCCGCCTC
OT003	chr11:72237613-72237822	GCTTCCCTCCACCCCGCATC
OT004	chr14:93059653-93059880	CACCCCCTCCACCCCTGCCT
OT005	chr3:45306869-45307048	TCCCCCCCCCCACCCCACCCC
OT006	chr3:49812695-49812863	GACCCCTACCACCCCATCTC
OT007	chr2:232480415-232480595	CCTCCCCTCCACCCCGCTCC
OT009	chr6:26470334-26470549	GACCCCCCCCCACCCCACCCC
OT012	chr14:74631890-74632068	CCTCACCCCCACCCCACCTC
OT013	chr11:12287317-12287554	ATCCCCCTCCACCCCACCCC
OT014	chr9:131282589-131282766	AACACCCTCCACTCCCCCTC
OT016	chr22:47115291-47115481	GCCCCCCACCACCCCACCCC
OT017	chr11:66851086-66851251	GGCAGCCTCCACCACGCCTC
OT018	chr17:4455366-4455527	TACCCCCCACACCCCGCCTC
OT019	chr1:109629350-109629530	TCCTCCTCCACCCACGCCTC
OT020	chr11:61553809-61554043	CACCCCCTCCCCTCCGCCTC
OT021	chr20:58197028-58197262	CTCCCACTCCACCCCGTCAC
OT022	chr3:14898799-14898971	AGTCCCCTCCACCCCTCCTC
OT023	chr20:63944421-63944611	CCGCCCCCCCCATCCCGCCCC
OT024	chr13:41515568-41515760	TGACCCCTCCACCCCGTCAC
OT025	chr1:1358817-1359030	TGGCCCCTCCGCCCCACCTC
OT026	chr17:42022714-42022907	AGTGGCCTCCACCCCACCTC
OT027	chr1:23435702-23435860	ACTCCCCTCCACCCCACCTC
OT028	chr2:128486464-128486697	CCCCCCCCCCCACCCCGCCCC
OT029	chrX:153571545-153571748	GTCCCCCTCCTCCCCACCTC
OT030	chr1:178769496-178769671	GGCCCTCTCCACTCCACCTC
OT031	chr1:117563396-117563579	CCCACCCTCACTCCCGCCTC
OT032	chr2:80967173-80967342	GGCCCCCTCCACTCCACCCC
OT033	chr12:57210411-57210629	GCCACCCACCACCCCACCTC
OT034	chr12:49881096-49881261	GCCCCCACCCACCCCGGCTC
OT035	chr17:60327450-60327624	CGCCCACCCCACCCCACCTC

Table B.6: VEGFA2 rhAmpSeq primer pool

Target	Chr. location	Guide sequence
OT036	chr2:130301559-130301734	TGCCCCCCCAACCCCACCTC
OT037	chr7:100688159-100688347	GGTCCCCTCCTCCCCACCCC
OT038	chr22:43288376-43288561	AGCCCCCACCTCCCCGCCTC
OT039	chr15:101159524-101159684	ACACCCTCCACCCACACCTC
OT040	chr11:31795806-31795990	GGGCCCCTCCACCCCGCCTC
OT041	chr12:105659835-105660018	TCCCCCTCCACCCCAGCCTC
OT042	chr9:136982167-136982354	GAGTGTCTCCACCCCGCATC
OT043	chr20:37362658-37362821	AGACCCCCCCACCCCACCCC
OT044	chr11:133907457-133907631	TCTCCCCTCCACCCCGCCCT
OT045	chr12:56239617-56239797	GGCCCCATCCACCCCACCTC
OT046	chr14:68671495-68671698	CTCCCCCAACACCCCGCCTC
OT047	chr22:27834988-27835181	AACACCCCCAGCCCCACCTC
OT048	chr13:114029878-114030087	CAGCCCCTCCACTCCACCCC
OT049	chr2:128319907-128320132	CTCCCCCCCCCACCCCGTCCC
OT050	chr19:13011318-13011498	GCCCCCCACCACCCCACCTC
OT051	chr10:652266-652467	CCCCCCCGCCACCCCACCCC
OT052	chr7:151023958-151024119	ATCCCACTCCACCCCACCCC
OT053	chr17:49577108-49577281	CCGTCCCTCTACCCCACCTC
OT054	chr8:143707498-143707691	GCTCACCCCCACCCCGCCCC
OT055	chr3:69929316-69929478	CTCCCACTCCACCCCACCCC
OT056	chr5:171619257-171619494	GGGCCCCCCCCACCCCACCCC
OT057	chr3:9501187-9501348	ATCCCCCAACTCCCCACCTC
OT059	chr20:10933245-10933463	CCACCCCCCCCACCCCGCCCC
OT060	chr17:41888402-41888586	TGCCCCTCCCACCCCGCCTC
OT061	chr22:44738920-44739100	CACCCACTACTCCCCGCCTC
OT062	chr14:18920658-18920897	GCCCCCCCCCAACCCCACCAC
OT064	chr9:100837183-100837427	ACACCCCCCCCACCCCGCCTC
OT065	chr16:75095812-75095976	ACCCCCCTACACCCCACCAC
OT066	chr9:129106894-129107070	TCACCCCTCCCCCTCGCCTC
OT067	chr14:93058391-93058574	CACCCCCTCCACCCCGCCTC
OT068	chr15:26911992-26912150	CCTCCCCTCCACCCCACCTC
OT069	chr8:18183937-18184109	CCCCCCCACCACCCCGCCCC
OT070	chr20:62428343-62428500	CACTCCCTCCTCCCCGCCTC
OT071	chr10:44064529-44064698	CACTCCCCACACCCCACCTC
OT072	chr3:194095670-194095895	TCCCCTCTCCTCCCCACCTC
OT073	chr21:46020453-46020696	AGACCCCCCCCACCCCACCCC
OT074	chr9:33202521-33202707	GTCCCCTCCCTCCCCGCCTC
OT075	chr2:54031101-54031341	CCCTCCCCCCCACCCCACATC
OT076	chr17:50547363-50547537	CCCTTCCCCCCACCCCACCTC
OT077	chr7:44213887-44214071	GGCCCACTCCACTCCGTCTC
OT079	chr2:240992656-240992826	GACCCCCTCCACCCACCATC
OT080	chr2:238153204-238153364	CACCCCCCCCCACCCCGTCTT

Table B.6: VEGFA2 rhAmpSeq primer pool

Target	Chr. location	Guide sequence
OT081	chr1:920833-921007	TCTCCTCCCCACCCCACCTC
OT082	chr2:169716746-169716913	CCACCCCCCCACCCCGCCCC
OT083	chr12:86287447-86287608	TAGCCCCTCCACCCCACCAC
OT084	chr11:134489926-134490101	ATCCCCTCCTACCCCGCCTT
OT085	chr16:70696946-70697155	CCACCCCTCCACTCAGCCTC
OT086	chr10:70778382-70778548	CAGTCCCCCCACCCCACCTC
OT087	chr11:44620077-44620235	ATCCTCCTCCTCCCCGCCCC
OT088	chr22:37655536-37655707	TCCCCCCTCAGCCCCGCCCC
OT089	chr5:80028128-80028291	TCCCCACCCCACCCCACCTC
OT090	chr5:102661158-102661317	TTCCCACTCCATCCCGCTTC
OT091	chr2:127416800-127417031	ATCCCCCCCCAACCCCACCCC
OT092	chr16:55822530-55822698	CACTCCCTCCTCCCCACCCC
OT093	chr19:48318566-48318750	GACCCCCTCCTCCCCAGCTC
OT094	chr5:149810305-149810545	TAGGCCCCCCACCCCACCTT
OT095	chr12:132197193-132197372	GTTCCCCCTCCCCCCACCTC
OT097	chr2:25343002-25343194	CACCCCCCCACCCCCGCCTC
OT098	chr10:124546065-124546220	CACCCCTTCCACCCCACCTG
OT099	chr1:112691640-112691814	ACTACCCACCTCCCCGCCTC
OT100	chr2:241275089-241275255	ATTCCCCCCCACCCCGCCTC
OT101	chr14:68770164-68770395	CTGTCCCCCCACCCAGCCTC
OT102	chr8:133165470-133165636	CCTCCCCTCCACCTCGCCTC
OT103	chr2:11134441-11134613	TTCCCCCTCCTCCCAGCCTC
OT104	chr12:25872069-25872237	CATTCCCCCCACCCCACCTC
OT105	chr17:64089585-64089773	ACTCCCCTCCACCCCGGCTC
OT106	chr2:11559727-11559893	CTCCCTCCCCACCCCACCTC
OT107	chr8:143740695-143740928	GTACCCCACCACCCCGCCCC
OT108	chr2:43381337-43381521	TACTCCCCCCACCCAGCCTC
OT109	chr12:3080696-3080865	TTCCCCCTCCTCCCAGCCTC
OT110	chr7:45178255-45178415	GCCCCACCCCACTCCACCTC
OT111	chr17:74826763-74826922	AATCCCCACTACCCCACCTC
OT112	chr1:42502600-42502820	TGCCCCCTCCCCCCAGCCTC
OT113	chr15:85285292-85285472	CTTTCCCTCCACCCAGCCTC
OT114	chr22:42995030-42995204	ATGTCCCCCCTCCCCGCCTC
OT115	chr18:1341088-1341311	CCATCCCTCCACCCCACATC
OT116	chr10:44155225-44155450	CACCCCACCCACCCCTCCTC
OT117	chr2:47551911-47552110	TCCCCCCCACCTCCCCACCAC
OT118	chr2:51769068-51769236	AACTCCCCACACCCCACCAC
OT119	chr8:143891897-143892100	CTTCCCCTTCACCCCAACTC
OT120	chr21:15809882-15810071	CTCCCCCTCCCCCCCAGCCTC
OT121	chr14:69014204-69014373	ACCACCCTCCGCCCCGCCTC
OT122	chr11:2664888-2665072	CTCACCCCCCCACCCCACCTC
OT123	chr10:110211236-110211426	TCCACCCCCCCACCCAGCCTC

Table B.6: VEGFA2 rhAmpSeq primer pool

Target	Chr. location	Guide sequence
OT124	chr4:3630980-3631215	CCCCCCTCCCACCCCACCCC
OT125	chr5:41556436-41556608	TTGCCCCTCCTCCCAGCCTC
OT126	chr2:26118286-26118511	CCCCACCCCCACCCCACCCC
OT127	chr3:13001929-13002127	GCCACCCCCCACCCTGCCTC
OT128	chr5:156894002-156894179	GACCCCACCTACCCCACCTC
OT129	chr1:9824725-9824924	GGCCCCCTCCTCCTCGCCTC
OT130	chr10:77721914-77722087	ACTCCCACCCACCCCACCTC
OT131	chr3:126406315-126406491	TCCCTCCCCCATCCCGCCCC
OT132	chr17:28344110-28344274	CATCCCCTC-ACCCCGCCTT
OT133	chr15:35866793-35866960	GTGACCCCCCACCCCACCCC
OT134	chr12:123409062-123409255	TTCCCCCCCCAACCCCACCTC
OT135	chr22:39490971-39491150	TGTCCCCTCCCCCCCACCTC
OT136	chr5:6714884-6715050	CTACCCCTCCACCCCGCCTC
OT137	chr7:105293382-105293601	TCCACCCCCCACCCCGCCCC
OT138	chr3:31417058-31417232	CTTCCCCCACACCCCGCCCC
OT139	chr9:27338819-27338992	GACCCCTCCCACCCCGACTC
OT140	chr22:50446229-50446418	CCCCCCCCCCCCCCCGCCTC
OT141	chr10:114534409-114534600	CCCCACCCCCACCCCGCCTC
OT142	chr18:23779514-23779688	GCCCCCACCCACCCCGCCTC
OT144	chr17:17051421-17051615	CTCCCCCGCCACCCCGCCCC
OT145	chr16:56929436-56929650	TGCCCCCCCCCACCCCACCTC
OT146	chr3:140679856-140680025	CAACCCCCCCCACCCCGCTTC
OT147	chr11:63623489-63623673	GACACCTTCCACCCCGTCTC
OT148	chr11:76784654-76784876	CACCCCCCCCCCCCCCACCTC
OT149	chr1:939889-940071	GACCCTGTCCACCCCACCTC
OT150	chr4:8840130-8840290	CATACCCCCCACCCCGCCCC
OT151	chr2:12604546-12604707	GACACACCCCACCCCACCTC
OT152	chr10:133336322-133336516	CGCCCTCCCCACCCCGCCTC
OT153	chr5:14457318-14457513	CCCCCCCCCCCGCCCCGCCCC

Table B.7: HEK4 rhAmpSeq primer pool

Target	Chr. location	Guide sequence
HEK4	chr20:32761907-32762073	GGCACTGCGGCTGGAGGTGG
OT001	chr11:528837-529011	GGCTCTGCGTCCGGAGGTGA
OT002	chr7:1357681-1357848	AGCACTGCAGCTGGGAGTGG
OT003	chr20:61435376-61435557	TGCACTGCGGCCGGAGGAGG
OT004	chr9:5020534-5020713	TGCACTGCAGCTGCAGGTGG
OT005	chr19:32891077-32891260	GGCTCTGCGGCTGGAGGGGG
OT006	chr8:144319943-144320109	AGCCCTGCGGCCGGGGGAGG
OT007	chr3:49017834-49017995	GGGACTGCGGCTGGAGGTGG
OT008	chr15:40751949-40752113	GGCGCTGCGGCGGGAGGTGG

Table B.7: HEK4 rhAmpSeq primer pool

Target	Chr. location	Guide sequence
OT009	chr15:75339695-75339868	GCACCTGCGGCTGGAGGTGG
OT010	chr17:75504962-75505135	GCACCTGCGGCCAGGGGTGG
OT011	chr1:201098144-201098317	GGCACTGCTGCTAGAGGTGC
OT012	chr10:44821939-44822104	GGCTCTGCAGCTGTAGGAGG
OT013	chr12:53059695-53059869	TGGACTGCGGCTGGAGAGGG
OT014	chr8:142477083-142477265	GCACTGCGTGCTGTAGGAGG
OT015	chr22:41223988-41224169	GGGCATGCGGCTGGAAGTGG
OT016	chr20:59341832-59342037	AGCACTGAGGCTG-AGGTGG
OT017	chr16:1022497-1022668	GGCCCTGCAGCAGGGGGTGG
OT018	chr14:21525161-21525391	GGTACAGCGGCTGGGGGAGG
OT019	chr20:62320509-62320740	GGCACAGCAGCTGGAGGTGC
OT020	chr8:143698954-143699195	GACACTGCAGCTGGAGGTGG
OT021	chr6:3277369-3277544	TGCACTGCGGCGGGAGGAGG
OT022	chr14:24270990-24271171	GGCACTGCCACTGGGGGTGA
OT023	chr11:65613352-65613529	GGAGCTGCGGCTGGAGGAGG
OT024	chr3:51691371-51691546	GGCTCTGTGGCTGGAGGAGG
OT025	chr6:160096805-160096990	GGCACTGCTGCTGGGGGTGG
OT026	chr20:1171140-1171301	GGCACTGTGGCTGCAGGTGG
OT027	chrX:105601969-105602132	AGCTCTGCGGCTGGAGGAGG
OT028	chr17:33463707-33463900	TGCACTGCAGCTGGGGGCAG
OT029	chr16:88930518-88930701	GACACCGCGGATGAAGGTGA
OT030	chr9:10651098-10651258	GGCCCTGCAGCTAGGGGTGT
OT031	chr1:955159-955345	GGCACTGTTGCTGGTGGTAG
OT032	chr1:154354587-154354776	GGCTCTGCGGCTGAAGGAAG
OT033	chr9:130276767-130277009	GTCACTGCAGCTGGAGGAGG
OT034	chr19:38125440-38125677	GGCACTGAGACTGGGGGTGG
OT035	chr10:75343253-75343412	GGCATCACGGCTGGAGGTGG
OT036	chr16:50300801-50300992	GGTTCTGCGGTTGGGGGTGG
OT037	chr10:13650488-13650711	GGCACTGGGGCTGGGGGAGG
OT038	chr8:892192-892380	TGCACAGCAGCCAGAGGTGG
OT039	chr2:240897411-240897628	GCCACTGCGGCCGAGGGTGA
OT040	chr8:22108189-22108379	GGCGCTGCAGCGGGAGGTGA
OT041	chr19:1295012-1295189	GACACTGAGGCAGGAGGTGG
OT042	chr22:48736969-48737152	AGCACAGCAGCTGCAGGTGG
OT043	chr4:55948933-55949085	GGCAATGCGGCTGGAGGCGG
OT045	chr8:143923203-143923425	AACACTGCGGCTGCAGCTGG
OT046	chr8:144052762-144052932	GGCACTG-GGCCGGAGGTGT
OT047	chrX:71616628-71616788	GGCCATGCGGCTGGTGGTGG
OT048	chr11:17517275-17517514	GCACTGCGGTCAGGAGGAGG
OT049	chr13:38688729-38688912	AGCAGTGCGGCTAGAGGTGG
OT050	chr7:139455-139691	TGCACCGCGGCTGGGGCTGG
OT051	chr2:151971349-151971530	GGCACTTCGGTTGGGGGTGG

Table B.7: HEK4 rhAmpSeq primer pool

Target	Chr. location	Guide sequence
OT052	chr7:128937601-128937761	AGCACTGCAGCGGGAGGTAC
OT053	chr1:111515631-111515814	TCTACTGCGGCTGGTGGAGG
OT054	chr22:22941777-22942016	GGCCCTGCTGCTGGAGGTGC
OT055	chr19:46383808-46383974	GAGGCTGCGGCTGGGGGTGG
OT057	chr11:10943838-10944065	CAGACTGCGGCTGGGGTAGG
OT058	chr19:15191616-15191781	GGCACTGCGGCCAGAGGGAG
OT059	chr12:113497503-113497705	GGCCCTGCGGCTGGAGATAT
OT060	chr3:165649225-165649398	CGCACTGCGGCTGCAGGTGG
OT061	chr10:1243119-1243339	GGCCCTTCGGCTGGAGGTGG
OT062	chr17:31488467-31488658	GGCGCTGCGGCCGGAGGTGG
OT063	chr6:149846000-149846157	CTCACTGCGGATGAAGGTGG
OT064	chr11:45105511-45105712	GTCACCACGGCTGGAGGGAG
OT065	chr6:16217187-16217424	TGCACTGGGGCTGCAGGTGG
OT066	chr7:139559596-139559773	GCCACTGCGACTGGAGGAGG
OT067	chr7:158502585-158502815	GACACAGTGGCTGGGGGTGA
OT068	chr6:158769799-158769978	GGCCCTGCAGCTGGAGGAGG
OT069	chr20:61505436-61505610	AGCACTGCAGATGGAGGAGG
OT070	chr4:176065566-176065737	TGCACTGAGGCTGAAAGTGG
OT071	chr12:54583769-54583997	GACACTGCCTCTGGGGGTGG
OT072	chr10:71675393-71675585	GTAACTGCGGCTGGCGGTGG
OT073	chr5:141853146-141853386	GGCACTGCGGCAGGGAGGAG
OT074	chr8:142481272-142481460	AACACAGCGGATGTAGGTGG
OT075	chr5:178249281-178249462	GCCACTGTGGCTGGAGGTGG
OT076	chr8:143925185-143925415	GGAGCTGCGGCTGCAGGCGG
OT077	chr4:6092326-6092520	GAGCCTGCGGCTGCAGGTGG
OT078	chr8:141353012-141353204	GGCACAGCAGCTGGGTGGGG
OT079	chr18:11866919-11867092	GGGACTGCGGCTGAGAGAGG
OT080	chr14:105472719-105472913	GGGACTGCGGCTGGGGTGAG
OT081	chr1:12199651-12199836	AGCACTGCAGCGGGAGGTGA
OT082	chr16:28255599-28255782	GGCTCTTCGGCTGGAGGTAG
OT083	chr9:130466253-130466448	GGCCCTGTGGCTGAAGGTGC
OT084	chr3:16774023-16774191	CGCACTGGGGCTGCAGGTGG
OT085	chr19:50716230-50716409	GGCATCGCGGCCGGAGGTGG
OT087	chr16:89402745-89402929	GGCACTGCGGGAGGAGGTGG
OT088	chr17:77433148-77433320	GACACCACGGCTGGAGATGG
OT089	chr20:63179323-63179548	GTCACTGCGGCTGCAGATGG
OT090	chr1:16929774-16929927	GGCTCTGCGGCTGGAGCCTG
OT091	chr17:9208927-9209088	GACTCTGCAGCTGAGAGTGG
OT092	chrY:13851584-13851767	AGCAGTGCGGCTGGGGGAGG
OT093	chr4:533494-533660	GCACTGCAGGCAGGGGGAGG
OT094	chr20:34999006-34999163	GGCACTGCGGCTGCAGGAGG
OT095	chr1:171049148-171049383	GCCACTGGGGCTGGGGGTGG

Table B.7: HEK4 rhAmpSeq primer pool

Target	Chr. location	Guide sequence
OT096	chr16:979851-980034	GGCACTGCAGACGGAGGTGT
OT097	chr5:642828-643053	GGCACAGCAGCTGGAAGTGG
OT098	chr17:75579111-75579278	AGCACTGCAGCTGG-GATGG
OT099	chr15:41511140-41511369	TGCACTGCGGGCGGAGGCGG
OT100	chr15:71394486-71394680	TGCTCTGCGGCAGGAGGAGG
OT101	chr7:54493604-54493807	AGGACTGCGGCTGGGGGTGG
OT102	chr7:110502974-110503141	GCCACTGCAGCTAGAGGTGG
OT103	chr19:40714512-40714684	GGCAATGTGGCTGAAGGTGG
OT104	chr16:50266357-50266553	AGCACTGTGGCTGGGGGAGG
OT105	chr13:88248631-88248792	CACACTGCAGCTGGAGGTGG
OT106	chr22:29734755-29734932	GGCTGTGCGGCCAGAGGTGG
OT107	chr7:135187151-135187373	AGCACTGTGGCTGGGGGAGG
OT108	chr17:326392-326574	TGCACTGTGGCTGGAGATGG
OT109	chr18:39614499-39614719	GGCACTGCGGGTGGAGGCGG
OT110	chr20:46714300-46714473	GGCACTGAGGGTGGAGGTGG
OT111	chr3:10377222-10377399	GGCTCCGCAGCTGGAGGTGG
OT112	chr2:45020160-45020324	GACACCGTGACTGGAGGTGG
OT113	chr8:118214780-118214956	GGCACAATGGCTGGAGGTGA
OT114	chr10:125006191-125006369	GGCACGACGGCTGGAGGTGG
OT115	chr13:27055206-27055386	GGCACTGGGGTTGGAGGTGG
OT116	chr14:88306587-88306805	AGCACTGGGGCTGGGGGAGG
OT117	chr20:38842614-38842793	AGCACTGTGCCTGGGGGTGG
OT118	chr16:88649638-88649797	AGCACGGCAGCTGGAGGAGG
OT120	chr17:74254895-74255087	GGTACTCTGGCTGGAGGTGG
OT122	chr17:40322061-40322270	GGCACCTTGGCTGAAGGTGG

C

Safety-, ethical and environmental disclosures

C.1 Safety aspects

C.1.1 General safety aspects

AstraZeneca follows the safety guidelines enforced by EU directive 90/219/EEG, the Swedish Environmental Code SFS 1998:808, and the Swedish provisions on the "Contained Use of Genetically modified Microorganisms" (AFS 2011:02 & SFS 2000:271). All researchers and personnel involved in this study have undergone basic safety and applied biohazard/biosafety training at AstraZeneca. There are additional documentation with safety guidelines for each laboratory and hazardous agent (Standard Operating Procedures - SOPs) and additional courses in the safe handling of potential chemical- and biohazards that is available- and mandatory to all personnel involved in such work. In line with AZ safety regulations, the Safety Data Sheet (SDS) from the vendor of each chemical and biological agent used in the experiments was closely studied prior to starting the experiments, and recommended safety precautions according to specific SOPs were taken during the handling and use of those chemicals and biological agents. The safety equipment in AZ's molecular biology and cell culture laboratories includes the use of standard laboratory safety equipment and personal protective equipment (PPE). This encompasses a combination of primary barriers such as laminar flow-biosafety cabinets and ventilated cabinets for handling of hazardous chemicals, enclosed containers for risk waste, and other engineering controls designed to remove or minimize exposure to hazardous materials, as well as personal protective equipment (PPE). PPE form an immediate barrier between the personnel and the hazardous agent, and they include gloves, protective eyeglasses, and laboratory coats, as well as shoe covers in the cell culture laboratories. Furthermore, since this project involved genetically modified human cells, which are classified as biorisklevel 2 (2L) according to AFS 2018:4, an L-notification to the Swedish Work Environment Authority was submitted and approved prior to the start of this project.

C.1.2 Hazardous chemicals and biological agents specific to this project

This study involved the use of a number of chemicals and biological reagents classified as potentially hazardous or in other ways a risk to the person handling them. Below is a list of all such reagents, their risk declaration, and any additional measures taken to minimize exposure and harm to the person handling them.

margin=2cm

C.1.3 Ethical aspects

Research in genome editing may fall under the Swedish Act concerning the Ethical Review of Research Involving Humans (2003:460) [26]. This includes research involving biological material derived from a living- or diseased human being, where the material can be traced back to that person. Such research must be approved by the Swedish Ethical Review Authority, which may only approve research that can be carried out with respect for human dignity and fundamental rights, and only if the risks that the research may pose for the health, safety, and privacy of the research subjects are out-weighed by its scientific value. The Act also contains provisions regarding information and informed consent. The Ethical Review Act does not apply to work carried out only as part of first-cycle (undergraduate) and second-cycle (Master's) courses and study programs. This exception from the ethical review is called the "student exception" (section 6.2 of the Government Bill 2018/19:165, approved by the Swedish Parliament). Thus, since this is a MSc Thesis Project it can be argued that it falls out of the scope of the Act.

This project could however be argued to fall under the European General Data Protection Regulation (GDPR) [85] since it involves the use of hiPSCs which have been derived from a human being and could potentially be traced back to that person through genome sequencing. Unlike human embryonic stem cells, human iPSCs usually originate from individuals from whom additional phenotypic, clinical, and behavioral data (such as family history, age of disease onset, medications, diagnostic results, etc.) may be accessed [76]. Unless the iPSC are derived from fetal tissue or a deceased donor, these individuals are 'natural or legal persons' and the data are 'personal' and 'health' data, which are considered especially sensitive and deserving of legal protection. All personal data are regulated by data protection law, meaning donors are entitled to a high degree of privacy protection and to the security of the data associated with the cell line [76]. Thus, at AZ, iPSCs fall under the category of Ethically Sensitive Human Material, which in turn is regulated under the AZ Global Standard and according to the AZ Global Standard on work with Human Biological Samples, the use of iPSCs is regulated by the AZ HBS Governance bodies. The project has therefore sought and received approval from this internal governing body prior to the start of experiments. The purpose of the AZ Global Standard is to ensure all research using Human Biological Samples (HBS) conducted by or on behalf of AstraZeneca (AZ) is both compliant with relevant legislation and regulations and shows appropriate sensitivity to global and local ethical and reputational concerns.

As for the gene editing performed in this study, it falls under the category of

somatic gene editing and has been approved by AstraZeneca, who in turn operates under aforementioned Swedish national legislation and overarching EU legislation.

C.1.4 Environmental aspects

As for environmental aspects of the project, the key concerns are high energy consumption from the use of laboratory equipment, the extensive use of single-use plastics (e.g. pipette tips and plastic tubes) and the use of potentially environmentally- and health-hazardous chemicals. These impacts cannot be avoided as each plays a crucial part in the experimental work. However, some measures can be taken to reduce the environmental impact of the project. Through the AZ Green Labs initiative, joint measures to improve the environmental sustainability of research facilities at AZ are promoted. This includes turning off machines after use (if possible), and planning the experimental work ahead to reduce unnecessary use of single-use plastics.

Furthermore, by ensuring that potentially biohazardous and chemically hazardous wastes are handled correctly (emptied in the chemical sink or collected in correctly labeled waste bins), toxic releases to the environment and to humans can be prevented. Finally, there is an ongoing transition within the Genome Engineering research facilities from the use of potentially health-hazardous chemicals, such as the DNA dye Ethidium bromide, to benign alternatives such as SYBR Safe.

Table C.1: List of potentially hazardous chemical and biological agents

Reagent	**Risk declaration**	**Reference**	**Additional safety measures**
β-mercaptoethanol	Acute toxicity (oral, inhalation & dermal), skin irritation/sensitization, serious eye damage, reproductive toxicity, specific target organ toxicity from repeated exposure (oral, liver & heart), short-term (acute) aquatic hazard and long-term (chronic) aquatic hazard	β-ME SDS (Sigma-Aldrich)	Prepare RLT buffer in a ventilated cabinet and perform the cell-lysis and first DNA/RNA extraction steps in a ventilated cabinet
Doxycycline (hyclate)	Acute oral toxicity, skin- and eye irritation, reproductive toxicity, specific target organ toxicity from a single exposure (respiratory system), and long-term (chronic) aquatic hazard.	Dox SDS (Sigma-Aldrich)	No
Qiagen AllPrep Mini/Micro Kit Reagents	Buffer AW1 contains guanidine hydrochloride, Buffer RLT Plus contains guanidine thiocyanate and Buffer RW1 contains a small amount of guanidine thiocyanate. Guanidine salts can form highly reactive compounds when combined with bleach.	*Qiagen AllPrep Mini Handbook* (Qiagen)	If liquid containing these buffers is spilled, clean it with suitable laboratory detergent and water. If the spilled liquid contains potentially infectious agents, clean the affected area first with laboratory detergent and water, and then with 1% (v/v) sodium hypochlorite.
STEMdiff™ SMADi Neural Induction Supplement	Causes skin irritation and serious eye irritation, combustible liquid	SMADi SDS (STEMCELL Technologies)	No

D

Supplementary results

D.1 Flow cytometry control samples

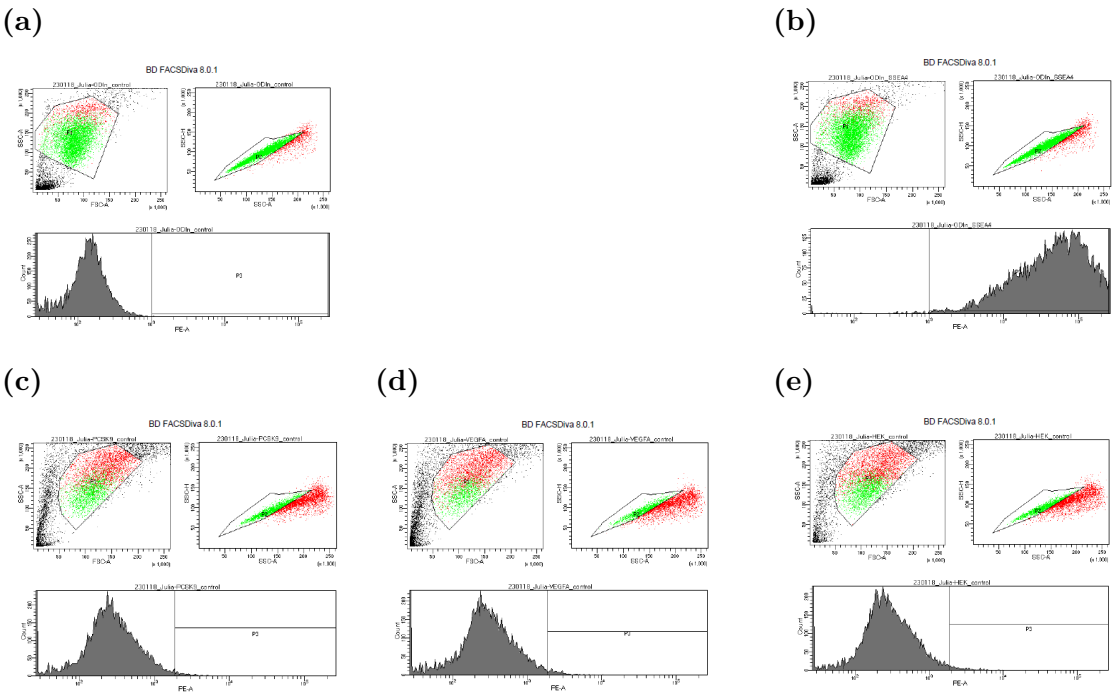


Figure D.1: Additional results from flow cytometry analysis of the three ODI-hiPSC lines for pluripotency validation with the pluripotency marker SSEA4 (SSEA4-PE). (a) isotype control (IgG) stained positive control ODIn-hiPSCs (without CRISPR-cassette), (b) SSEA4 stained positive control ODIn-hiPSCs (without CRISPR-cassette), (c) isotype control (IgG) stained ODIn-U6-PCSK9 hiPSCs, (d) isotype control (IgG) stained ODIn-U6-VEGFA2 hiPSCs, and (e) isotype control (IgG) stained ODIn-U6-HEK4 hiPSCs.

(a) **(b)**

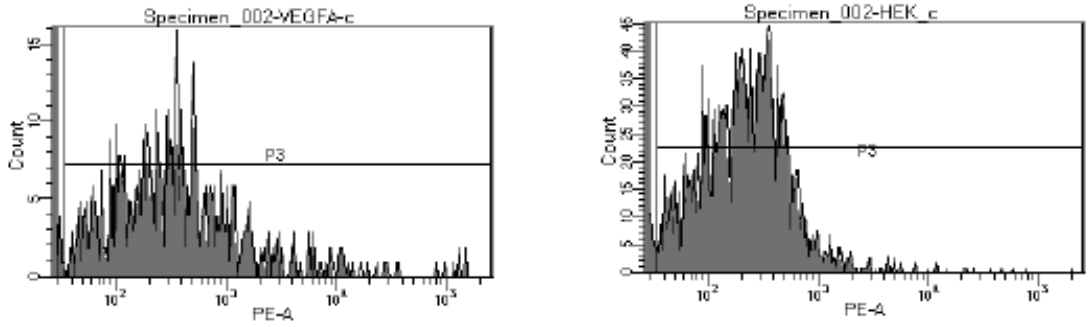


Figure D.2: Additional results from flow cytometry analysis of the first batch of the three ODIn-hiPSC-derived cardiomyocytes for characterization of the proportion of mature cells in the total cell population (one biological replicate per stain). All cell lines were stained with Cardiac Troponin T (cTnT) and an isotype control stain (IgG-PE), except for ODIn-PCSK9 (no IgG stain) due to low cell count at harvest. (a) isotype control (IgG) stained ODIn-U6-VEGFA2, and (b) isotype control (IgG) stained ODIn-U6-HEK4 cardiomyocytes. The subpopulation of cTnT-positive cells was gated based on forward- and side scatter (FSC/SSC), removing cell duplets and smaller particles.

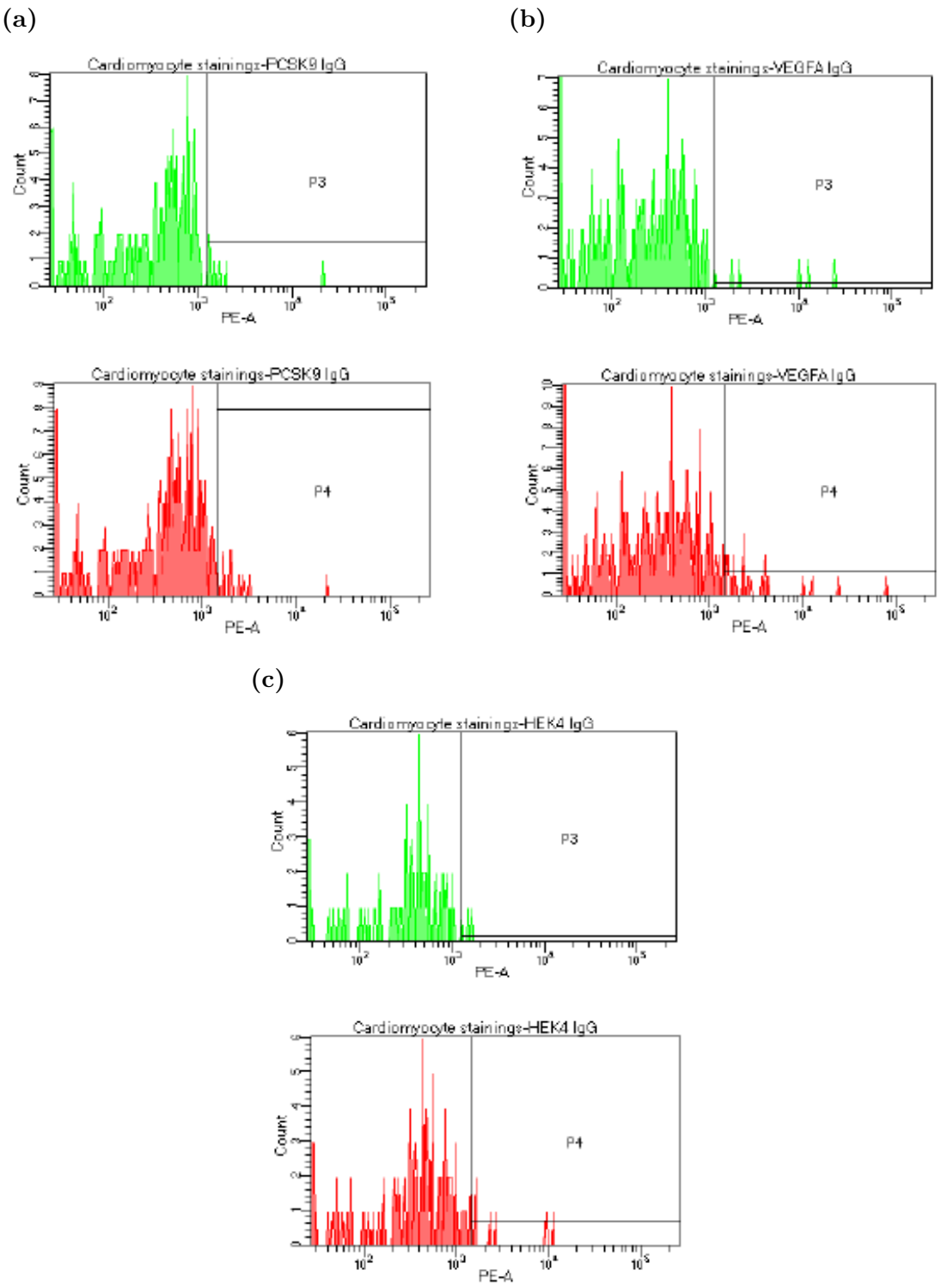


Figure D.3: Here, the additional results of the acquisition of IgG-PE and SSEA4-PE-stained CMs (batch II) are depicted, with gating based on SSC and FSC. cTnT stained positive control U2-OS cells failed. (a) isotype control (IgG) stained ODIn-U6-PCSK9, (b) isotype control (IgG) stained ODIn-U6-VEGFA2 CM, (c) SSEA4 stained ODIn-U6-VEGFA2, (d) isotype control (IgG) stained ODIn-U6-HEK4, and (e) SSEA4 stained ODIn-U6-HEK4 CM.

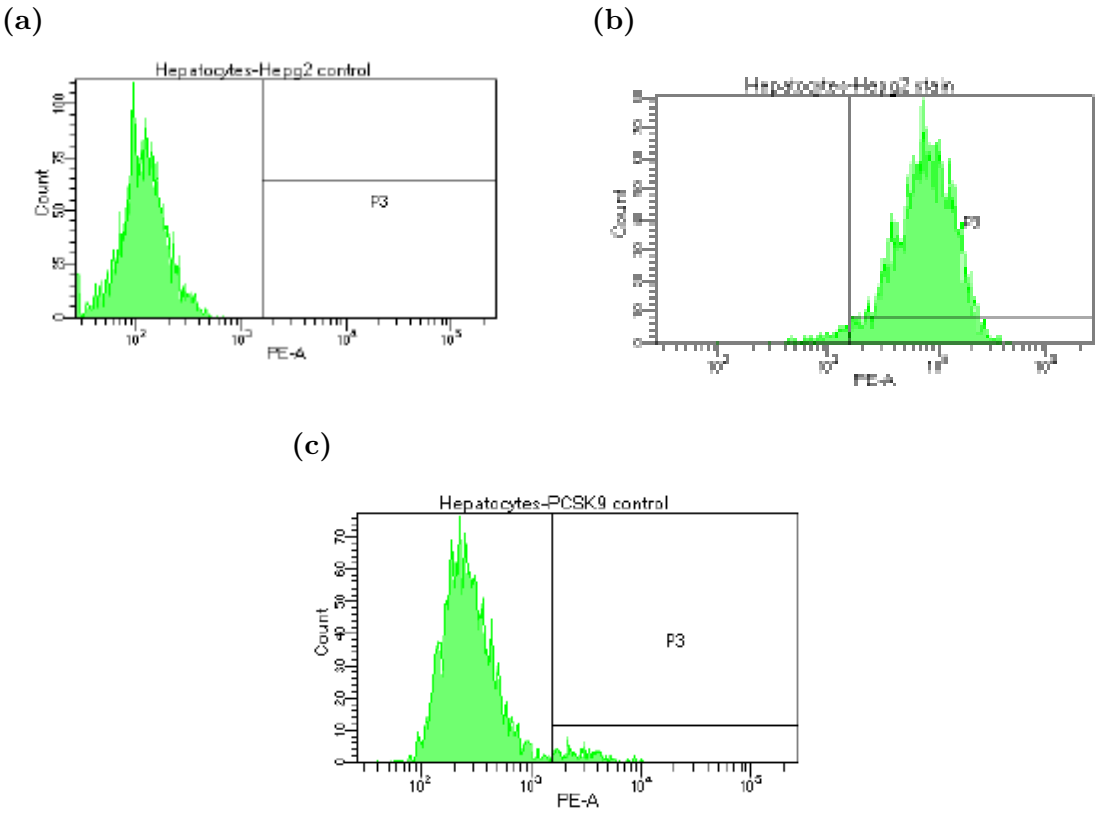


Figure D.4: Depicted here are the additional results from flow cytometry analysis of the two surviving cell lines of ODIn-hiPSC-derived hepatocytes with the Cellartis®Takara Bio kit. Both cell lines were stained with a PE-conjugated antibody targeting asialoglycoprotein receptor 1 (ASGPR1), but only ODIn-PCSK9 was stained for the isotype control (IgG-PE) and the pluripotency marker SSEA4 (SSEA4-PE). One biological replicate per stain was assayed and gating was based on SSC and FSC, removing any duplets and smaller particles. (a) isotype control stained positive control HepG2 cells, (b) ASGPR1 stained positive control HepG2 cells, and (c) isotype control (IgG) stained ODIn-U6-PCSK9. The hepatocellular carcinoma cell line HepG2 was assayed as a positive control for the expression of ASGPR1 and tested close to 100% ASGPR1-positive (96%).

(a) **(b)**

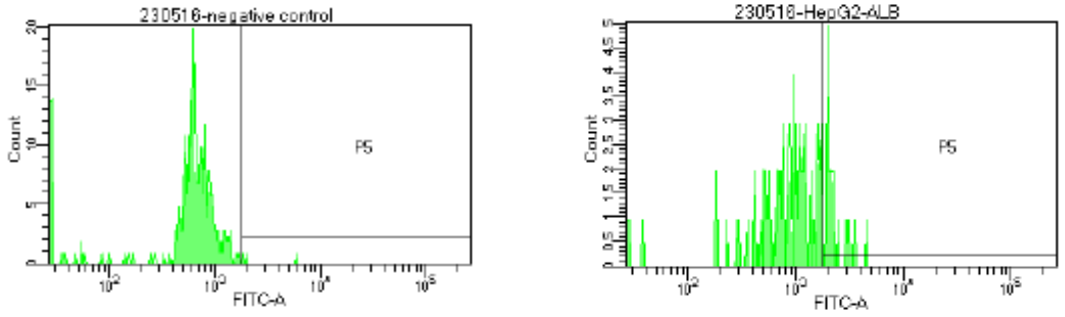


Figure D.5: Flow cytometry analysis of the second batch of ODIn-hiPSC-derived hepatocytes (STEMCELL Tech. kit). These hepatocytes were stained with PE-conjugated albumin (ALB-PE) instead of asialoglycoprotein receptor 1 (ASGPR1-PE), as well as an isotype control stain (IgG-PE) for the HepG2 control cell line (exclusively). This batch of ODIn hepatocytes was not stained for the pluripotency marker SSEA4 (SSEA4-PE), due to low cell count at harvest. Albumin-positive cells were gated based on SSC and FSC, removing any duplets and smaller particles. (a) isotype control stained positive control HepG2 cells, and (b) ALB-stained positive control HepG2 cells. The hepatocellular carcinoma cell line HepG2 was assayed as a positive control for the expression of albumin, as well as for the isotype control stain (IgG), but only 20% of HepG2 cells stained positive for albumin.

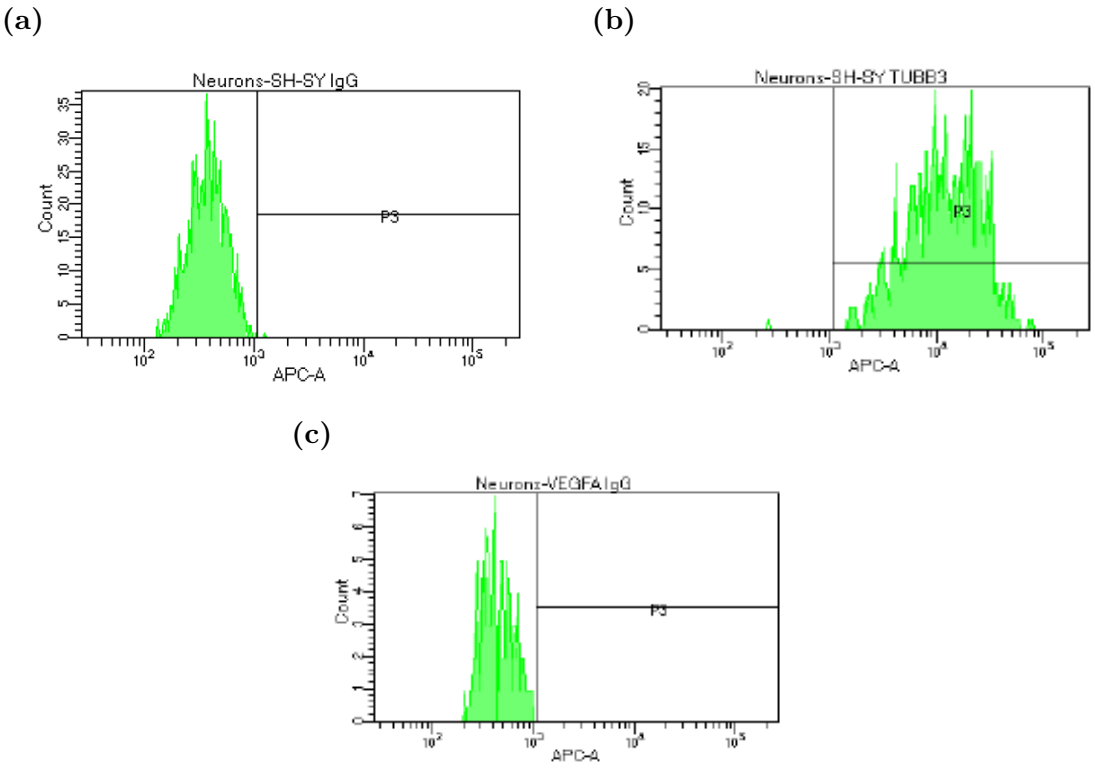


Figure D.6: Additional results from flow cytometry analysis of the two remaining cell lines of ODIn-hiPSC-derived neurons (batch I). Both cell lines were stained for APC-conjugated beta-tubulin class III (TUBB3), but only ODIn-VEGFA2 was assessed to have enough cells for an isotype control stain (IgG-PE). Neither were stained for the pluripotency marker SSEA4. TUBB3-positive cells were gated based on SSC and FSC, removing any duplets and smaller particles. (a) isotype control (IgG) stained positive control SH-SY5Y cells, (b) TUBB3 stained SH-SY5Y cells, and (c) isotype control (IgG) stained ODIn-U6-VEGFA. The neuroblastoma cell line SH-SY5Y was assayed as a positive control for the expression of TUBB3 and tested close to 100% TUBB3-positive (99.9%).

D.2 Results of the genotyping assays in ODIn neurons

The results from the initial PCR assay for genotyping the ODIn neurons were inconclusive, as significant smearing was confounding three of the wells and most fragments at the expected size were almost indistinguishable, see Figure D.7.

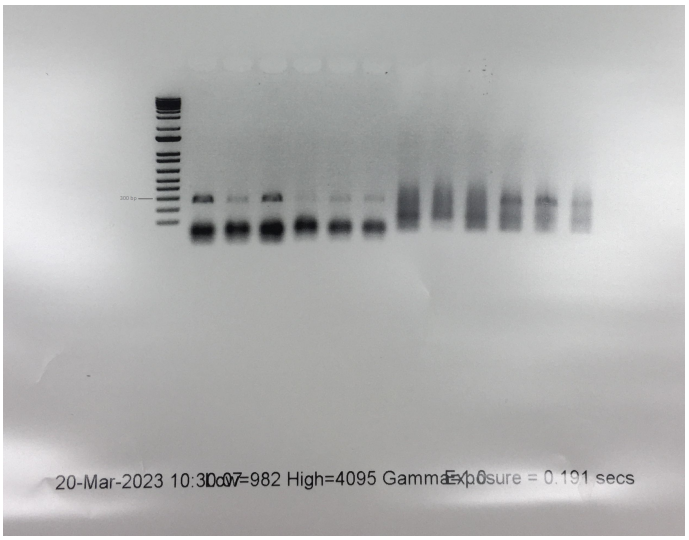


Figure D.7: Gel image after UV-exposure for 0.19 s, samples are in the order S1-S12 from left to right. The indicated fragment length is 300 bp and applies to both *PCSK9* and *VEGFA2* target sequences (approximately)

The second assay applied to genotype the ODIn neurons was Sanger-sequencing with TIDE analysis. The results from this assay were also inconclusive in terms of potential cell line mix-up. Some low-frequent editing was present in the ODIn-U6-VEGFA2 neurons at the *PCSK9* locus, indicating a potential mixed culture. No editing was detected at the VEGFA locus for ODIn-U6-PCSK9 cells, indicating an isogenic culture.

Finally, with targeted deep sequencing analysis (AmpSeq) of both target loci in both cell cultures of mature dopaminergic neurons, it could be confirmed that editing had indeed occurred at the *PCSK9* locus in both ODIn-U6-PCSK9 specimens and ODIn-U6-VEGFA2 neurons, but not vice versa. Furthermore, the ODIn-U6-VEGFA2 culture revealed both induced and leakage-mediated editing at the *PCSK9* locus. See Figure D.8 for a visual representation of the targeted sequencing results. Consequently, the editing outcomes were interpreted as having (accidentally) generated one mixed culture of both cell lines ("ODIn-U6-VEGFA2"), whereas the ODIn-U6-PCSK9 neurons appeared to be an isogenic culture.

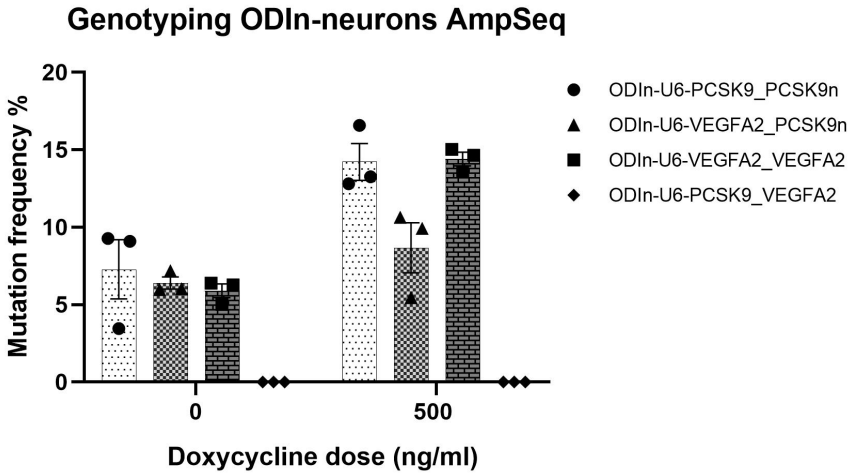


Figure D.8: Genotyping of the mature ODIn neurons with AmpSeq. Both *PCSK9* and VEGFA site-2 loci were sequenced in each respective neuron cell lineage.

DEPARTMENT OF BIOLOGY & BIOLOGICAL ENGINEERING
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY