Advanced CRISPR-Cas9 techniques for modulation of non-coding disease-associated genetic variants

Annika Sem Sippel Krill

This thesis is submitted in partial fulfillment of the requirements for the degree of Master of Science

Department of Biological Sciences

&

The Hormone Laboratory Research Group

Department of Clinical Science

University of Bergen

Norway

June 2023

Acknowledgements

The work presented in this thesis was conducted at the Hormone Laboratory at the Department of Clinical Science (Haukeland University Hospital), through the Department of Biological Sciences at the University of Bergen, from July 2022 to June 2023.

Much of this work has been performed in solitude in the laboratory, and during this time I have had the opportunity to reflect upon all of the incredible accomplishments science has achieved thus far – ranging from Nobel Prize-worthy discoveries to the development of everyday laboratory techniques. I am constantly amazed by how far the field as a whole has come, and the ingenious solutions invented to detect and comprehend that which is too small to see. Science is a collaborative endeavor, and no work is done in isolation. We all rely upon the contributions of those who have come before us, and I am particularly grateful to all those who have come before me and devoted their time and energy to this undertaking. The present moment feels like the perfect time to be a part of this journey; so much has been discovered and understood, and yet so much remains a mystery. It is an honor to be able to contribute to this collective human understanding of life and the universe, however small my contribution may be.

More specifically, I would like to thank my main supervisor Jan-Inge Bjune, for all the support and guidance you have given me. You have shared your time, experience, and knowledge with me, and I feel incredibly fortunate to have had you as a supervisor during this critical period of my life. In an age where time is the most valuable resource we have, you have shared yours most generously.

I would also like to thank my co-supervisor Simon Dankel, for valuable scientific discussion and guidance on the project and in writing this thesis.

Thanks to everyone in the Hormone Laboratory research group, for being so welcoming and supportive. Special thanks to Laura Roxana Jonassen, Mona Synnøve Bjune, Regine Åsen Jersin, Linn Skartveit, and Laura Lage Segura for teaching me specific laboratory techniques. Thanks to Nina Tangen Samulesen and Shanu Mirza, whom I have never met, but whose work I have built upon.

Finally, I would like to express my gratitude to my partner, Espen, for being such a warm, loving, and inspirational human being, as well as to my friends and family for emotional support and encouragement.

Bergen, June 2023

Annika Sem Sippel Krill

Contents

List of Abbreviations

AD-MSC	Adipose-derived mesenchymal stem cells
BMI	Body mass index
CA	CRISPR/Cas9-mediated enhancer activation
CA/I	CRISPR/Cas9-mediated enhancer modulation
Cas9n	Cas9 nickase
CI	CRISPR/Cas9-mediated enhancer inactivation
CRISPR	Clustered regularly interspaced palindromic repeats
CRM	*cis*-regulatory module
CVD	Cardiovascular disease
dCas9	Catalytically deactivated Cas9
dsDNA	Double-stranded DNA
eCas9	Enhanced Cas9 (higher target specificity)
eQTL	Expression quantitative loci
gDNA	Genomic DNA
GFP	Green fluorescent protein
GG	Golden Gate
gRNA	Cas9 guide RNA (all categories)
GWAS	Genome-wide association studies
HDR	Homology-directed repair
IFU	Infectious unit / lentiviral particle capable of transduction
LD	Linkage disequilibrium
MOI	Multiplicity of infection
MSC	Mesenchymal stem cells
ngRNA	Nicking guide RNA (for the PE3/PE3b system)
NHEJ	Non-homologous end-joining
PAM	Protospacer-adjacent motif
pDNA	Plasmid DNA
PE	Prime editing
PE2	Prime editor 2
PE3b	Prime editor system 3b
pegRNA	Prime editing guide RNA
PMCA	Phylogenetic module complexity analysis
PP	Physical (lentiviral) particle
RT	Reverse transcriptase
RT	Room temperature
SAT	Subcutaneous adipose tissue
sgRNA	Single guide RNA
SN	Supernatant
SNP	Single nucleotide polymorphism
ssDNA	Single-stranded DNA
T2D	Type 2 Diabetes
TF	Transcription factor

VAT	Visceral adipose tissue
Cas9	SpCas9 (Cas9 derived from *Streptococcus pyogenes*)

Summary

The search for genetic explanations of individual susceptibility to complex polygenic diseases has greatly intensified in recent years, with GWAS studies having successfully linked several hundred thousand genomic loci to complex diseases. However, the disease mechanisms underlying these associations remain largely unknown due to several limitations of GWAS. In particular, since 90 % of identified disease-associated SNPs are located outside of protein-coding regions, their biological effects are largely unclear, and although they likely affect gene regulatory functions via altered DNA motifs for specific transcription factors, the target genes need to be identified. Furthermore, each locus contains up to several hundred SNPs in linkage disequilibrium, and thus the causal SNPs are unknown.

Recent progress in addressing the problem has been made through the development of a more systematic approach involving several bioinformatic and experimental advances, that specifically tackles the mechanistic limitations of GWAS. The approach applies a series of methods to systematically narrow down the number of candidate causal SNPs, and ultimately identify the causal SNP and affected cell types, enhancers and gene regulatory mechanisms.. To experimentally validate causal SNPs *in cellulo* and establish downstream target genes and phenotypes, genome editing of the causal SNP must be performed. While regular CRISPR/Cas9 theoretically can be used for this purpose, it is highly inefficient and introduces several issues such as double-stranded breaks and potential off-target effects. In contrast, the newly developed Prime Editing (PE) technology may prove to be ideal for this type of precise genome editing. Furthermore, a second method, CRISPR/Cas9-mediated enhancer modulation (CA/I), may be used to strengthen the findings through direct epigenetic activation or repression of the enhancer in which the SNP resides.

As a proof-of-concept study of these recent advances, this thesis builds upon previous unpublished work from our group, which identified a likely causal SNP (rs1799993) in an enhancer associated with visceral obesity, a particularly harmful type of fat accumulation. Epigenetic data suggested that the SNP affects an enhancer element that is active in adipose-derived mesenchymal stem cells (AD-MSCs). Thus, the current study has focused on establishing the genome editing tool PE for in *situ editing* of the SNP, as well as the epigenetic modulation system CA/I for modulation of the surrounding predicted enhancer element, for use in AD-MSCs.

Spacers for sgRNAs and pegRNAs (the latter in PE) targeting the enhancer region and the SNP, respectively, were designed by *in silico* analysis, and evaluated *in vitro* for on-target efficiency. Extensions for pegRNAs were designed for editing the SNP from risk to protective allele and vice versa, and a plasmid library of the sgRNAs and pegRNAs was then prepared and sequence-validated. Furthermore, an appropriate mesenchymal stem cell (MSC) model was obtained and genotyped for the SNP in question. Because the MSCs are notoriously hard to transfect, comprehensive testing of transfection methods in these cells was performed, including a variety of chemical and physical methods of gene delivery.

While the required plasmids for both PE and CA/I were successfully made, no transfection method proved successful in MSCs using the large-size PE plasmids. Consequently, the SNP rs1799993 was not edited. However, nucleofection was identified as the method that gave the best results in this cell type using smaller plasmids, thus suggesting that optimizations should be directed toward reducing the PE-plasmid size for successful use of this method is MSCs. For similar reasons, a pilot lentiviral transduction of plasmids for CRISPR activation/repression did not result in stably transduced MSCs.

In summary, the work of this thesis has laid the groundwork for utilizing the PE and CA/I methods as tools to help translate GWAS association signals into causal gene regulatory mechanisms. Once optimized, the techniques should be able to determine whether rs1799993 is a causal SNP in the visceral obesity associated locus of interest, as well as identify the target genes of the enhancer where the SNP resides and alters transcription factor binding.

1. Introduction

1.1. The challenges of linking individual genetic variation to complex traits

Ever since the discovery of the principles of genetic inheritance, tremendous scientific efforts have been made towards uncovering the link between phenotypic traits and the underlying genetic variation (Claussnitzer et al., 2020). In particular, uncovering the genetic polymorphisms that contribute to disease and elucidating the mechanisms involved, holds the potential to not only develop targeted therapies for affected individuals, but may also revolutionize our understanding of human biological function (Sinnott-Armstrong et al., 2021). Indeed, studying monogenic diseases has greatly contributed to our mechanistic understanding of pathophysiology on the molecular level, including the genes, proteins, and pathways involved (Antonarakis & Beckmann, 2006). These diseases are generally easiest to discover, as they are caused by mutation(s) in a single gene, usually affecting protein function, and tend to have a recessive or dominant mode of inheritance with high penetrance (Lappalainen & MacArthur, 2021). The diseases typically have a qualitative phenotype with symptoms that become apparent before reproductive age, and thus the disease-causing alleles are infrequent in the population due to naturally being selected against. However, due to large effect size and the direct cause-effect relationship between gene and cellular phenotype, this type of mutation has provided invaluable insights into protein function and pathway interactions (Antonarakis & Beckmann, 2006; Hardies et al., 2016). In contrast, when it comes to more complex, polygenic diseases – despite decades of intense work – the progress has been disappointingly slow (Barroso & McCarthy, 2019).

Polygenic diseases are influenced by tens, hundreds, or even thousands of genetic variants, each with a small individual effect size, but that when combined can promote disease (Uffelmann et al., 2021; Yashin et al., 2012). The diseases are common, with an underlying quantitative trait (e.g. body adiposity, cholesterol levels), which usually does not progress to disease until later in life (e.g. Type 2 diabetes [T2D] or cardiovascular disease [CVD]) (Locke et al., 2015). Disease progression also depends on gene variant-interaction with environmental and lifestyle factors, and are therefore also referred to as complex diseases or multifactorial inheritance diseases (Ahmad et al., 2013; Manuck & McCaffery, 2014).

The complex nature of polygenic diseases has resulted in the task of identifying the contributing genetic variants to be exceptionally challenging (Uffelmann & Posthuma, 2021). Moreover, the causal alleles are relatively common in the population, and this, combined with their small individual effect sizes, means that the signal-to-noise ratio is low; and the weaker the signal, the larger the sample sizes required to detect it (Visscher et al., 2017). Furthermore, additional difficulties arise from the fact that the majority of the suspected variants are located outside of protein-coding DNA regions (Maurano et al., 2012), and thus, once the genetic variants have been detected, determining their genetic functions and individual mechanistic contributions to disease may be an equally daunting task (Uffelmann & Posthuma, 2021). However, in the course of the past 25 years, a series of catalyzing accomplishments in the field of genetics, as well as advances in technology

and computational power, have resulted in the development of new strategies that may allow us to begin uncovering some of these connections (Claussnitzer et al., 2020).

1.1.1. The reference genome and common genetic variation

Perhaps the most important achievement in the quest for understanding human genetics since the discovery of the DNA double helix, was the completion of the Human Genome Project (HGP) in 2003 (Hood & Rowen, 2013). The project, which sequenced the entire human genome for the first time, has revolutionized the fields of genetics and cell biology, both through lessons learned directly, as well as the methodological developments and new research approaches that it enabled (Claussnitzer et al., 2020). At the time, one of the most surprising (and perhaps disappointing) revelations was that the human genome contained only a fraction of the 50-140 000 genes originally predicted (Venter et al., 2001). Of the over 3 billion long nucleotide sequence, a mere 2 % were found to be protein-coding, encompassing only ~20 - 25 000 genes (International Human Genome Sequencing, 2004). Considering that the simple nematode *Caenorhabditis elegans (C. elegans)* had recently been revealed to harbor 19 000 genes (C elegans Sequencing Consortium, 1998), the established number for humans seemed too low to explain our much more complex biology (Venter et al., 2001). In light of this revelation, it became apparent that the answer likely resided in the other 98 % of the genome, with one possible explanation being complex intricate network of gene regulatory mechanism (Venter et al., 2001).

Following the successful sequencing of the human genome, the next large-scale collaborative project was to characterize the genetic variation found within and across populations. No two (non-identical) individuals have the exact same DNA sequence, and the average difference between two unrelated individuals is about 1 in 1000 base pairs (bp) (1000 Genomes Project Consortium et al., 2015). The majority of this variation comes in the form of single nucleotide polymorphisms (SNPs), as well as small deletions or insertions (1000 Genomes Project Consortium et al., 2015). The mapping of polymorphic loci (DNA locations where two sequence variations (alleles) are circulating in the population) was undertaken first by The International HapMap Project (2002-2005), and subsequently by The 1000 Genomes Project (2007-2015) (1000 Genomes Project Consortium et al., 2015; International HapMap, 2005). Collectively, the two projects have identified about 20 million SNPs where the minor allele frequency (MAF) is above 0.5%, and 8-10 million of which have frequencies of 5-50% and are thus considered common (1000 Genomes Project Consortium et al., 2015). The average genome contains 4-5 million minor SNP alleles, while the remaining SNP loci match the reference sequence (1000 Genomes Project Consortium et al., 2015).

1.1.2. LD-blocks and haplotyping

The HapMap project was based on the realization that linkage disequilibrium (LD) is not purely a function of distance, but instead that genetic recombination tends to occur in designated DNA regions

(recombination hotspots), while being virtually absent in others (LD-blocks) (Claussnitzer et al., 2020; International HapMap, 2003; Slatkin, 2008). Thus, all SNP alleles within an LD-block tend to be passed on as a single unit (International HapMap, 2003). An LD-block can range from 10 – 100 kb in length, containing up to several hundred SNPs (Anderson & Novembre, 2003; Karlsson et al., 2019). The combined genotypes of all SNP alleles in an LD-block make up its haplotype, and in this context LD-blocks are often referred to as "haplotype blocks" or simply "haploblocks" (International HapMap, 2003).

The mapping of SNPs in the population found that humans have low haplotype diversity; for many LD-block, only two or three major haplotypes have been discovered (International HapMap, 2003, 2005). Thus, once the predominant haplotypes in the population had been mapped, it became possible to genotype just a few carefully chosen SNPs (tag-SNPs) and impute the genotypes of the remaining SNPs in that block (Figure 1.1) (International HapMap, 2003; Wallace et al., 2007). This principle reduced the number of SNPs necessary to screen in order to determine the bulk of an individual's genetic variation from over 10 million to as little as 200 000 (International HapMap, 2003). Consequently, it had suddenly become feasible to screen the entire genomes of large groups of individuals, in search of genetic variants associated with disease.

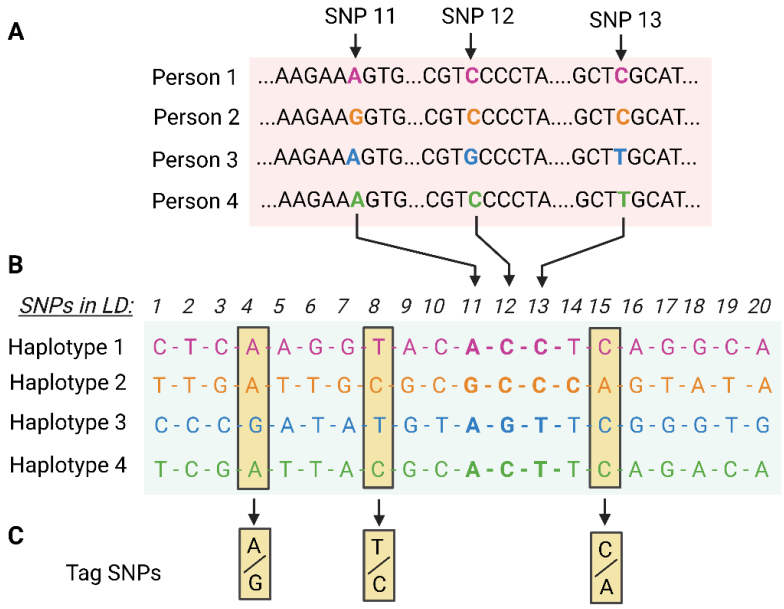


Figure 1.1 | From SNPs to haplotype to tag-SNPs. The human genome contains tens of millions of SNPs, but the fact that they are inherited in haploblocks mean that it is possible to only genotype a fraction of tag-SNPs and impute the remaining. **A)** The majority (99.9%) of the genome is identical between all humans, but variation occurs in numerous single nucleotide locations. Shown here are SNPs nr 11, 12, an 13 of an LD-block. **B)** Each SNP has two alleles, and the specific sequence of alleles an individual harbors in an LD-block make up the haplotype of that block. Humans have low haplotype diversity, and only a small number of haplotypes are common for any region. C) Since little to no recombination occurs within the haploblock, genotyping a couple well-chosen tag-SNPs is sufficient to determine the haplotype and thus impute the remaining SNPs of that block. *Figure made in Biorender, adapted from: (International HapMap, 2003).*

1.1.3. Genome-wide association studies – linking phenotype to genotype?

The results of the HapMap project, combined with improved high-throughput sequencing technologies, thus finally enabled the long-sought goal of performing genome-wide association studies (GWAS) on complex diseases in search of contributing genetic variants (Visscher et al., 2017). Since its inception in 2005, the

GWAS has become the primary approach to detect the underlying genetic variations that contribute to complex traits (Visscher et al., 2017). To examine qualitative traits or disease (e.g. T2D), the genomes of two groups of people are genotyped; one group of individuals that have the disease, and the other without (Uffelmann et al., 2021). SNPs that display uneven allele frequency between the two groups are said to be associated with the disease. For continuous traits, the trend of a trait within the population (e.g. BMI or fasting insulin levels) is compared to the trends in SNP allele frequencies (Uffelmann et al., 2021). The approach has so far associated over 200 000 loci with more than 3000 complex traits and diseases (Alsheikh et al., 2022).

Despite such a large number of associations, however, a surprisingly low number of SNPs have actually been linked to disease mechanism (Herman & Rosen, 2015; Lappalainen & MacArthur, 2021). Although the presence of LD-blocks has been essential for enabling GWAS studies with our current level of computational and sequencing power, it has equally left us with the challenge of untangling the LD SNPs from one another to identify the causal variant(s). And, as previously alluded to, the problem is further complicated by the fact that the majority of SNPs contributing to complex disease are located outside of protein-coding regions, and thus do not directly affect gene *function*, but more likely act indirectly through altered gene *expression* (Lappalainen & MacArthur, 2021). SNPs that lie within protein-coding sequences are generally straight-forward (though not necessarily easy) to link to disease mechanism (Alsheikh et al., 2022). First of all, the location of the SNP/locus instantly identifies the affected gene. And second, as the alternative SNP allele usually alters amino acid sequence or causes truncation of the protein product due to introduction of an early stop codon, the causal SNP is relatively easy to identify, and the disease mechanism can subsequently be elucidated by studying protein function and determining how it is affected by the mutation (Glazier et al., 2002; Thomas & Kejariwal, 2004). However, when it comes to SNPs in regulatory regions, the picture is markedly more complex.

The most abundant gene-regulatory element is the transcriptional enhancer. These *cis*-regulatory elements (CRMs; located on the same chromosome as the target gene) can regulate gene expression from thousands, or even millions, of bp away from the target gene(s) through the binding of transcription factors (TFs) (Carullo & Day, 2019). Enhancers are composed of multiple TF recognition sites that bind TFs in a cooperative manner, such that the presence of one TF enhances the binding of additional proteins to the adjacent sites (Carullo & Day, 2019). The TFs then recruit additional proteins (e.g. CTCF and Cohesin) that loop the DNA around to position the enhancer in close spatial proximity to the target promoter, and the multimeric TF complex bound to the enhancer can initiate gene transcription by recruiting and/or activating RNA polymerase II at the promoter (Figure 1.2)(Carullo & Day, 2019). Downstream gene expression is regulated by the exact combination of TFs at the enhancer, and the presence of an alternative SNP allele within an enhancer may disrupt the binding site of an essential TF, or even create a new binding site for a different TF, thus potentially altering downstream gene expression (Claringbould & Zaugg, 2021). The majority of enhancers have not been identified and their target genes are unknown, and thus identifying SNPs affecting this type of gene regulation

is thought to be one of the main bottlenecks of linking GWAS association to mechanistic function (Dey et al., 2022). Furthermore, enhancer elements are activated based on cell type, as they regulate tissue-specific gene expression (Carullo & Day, 2019), but since the DNA sequence does not vary between cell types, it holds no information about where or when the enhancer is active.

The problem thus amounts to the fact that a detected GWAS signal only points to a genetic locus being associated with disease, and does not alone provide information about target gene(s) or downstream biological pathways implicated in the disease. Moreover, because GWAS does not inform of which of the SNPs in the LD-block is the causal SNP, upstream mechanisms cannot be deduced, since the identity of the affected TF binding site is unknown. Finally, the cell type or tissue in which the SNP is active, as well as *when* it is active (i.e. during development or in the adult organism, and in response to which stimuli) is not known either. So despite the vast amount of GWAS data accumulated over the last two decades, the next major challenge of genetics is to devise systematic and effective methods that allow us to connect these associations to disease mechanism.

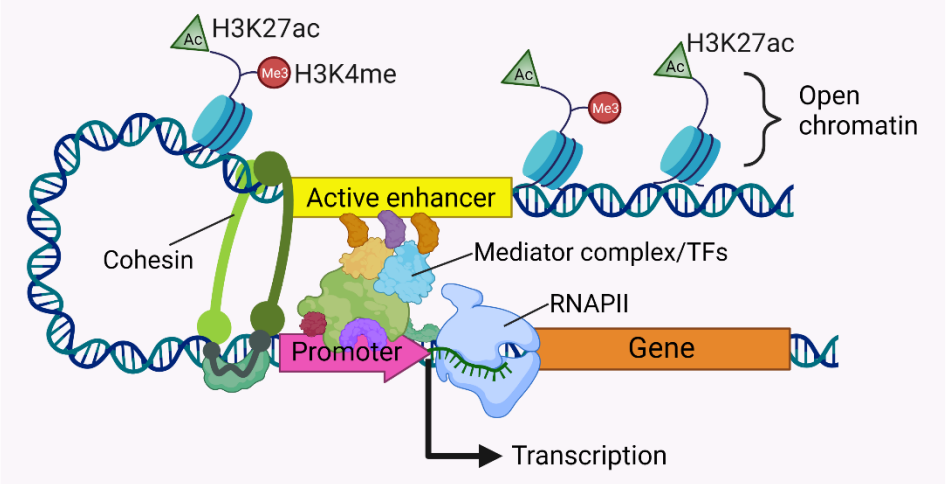


Figure 1.2 | Eukaryotic enhancer-activated gene transcription. Enhancers can regulate gene expression across long genomic distances, through the binding of TFs and recruitment of proteins (e.g Cohesin) that loop the DNA in a specific manner to place enhancer in close proximity to the promoter. The TFs bound at the enhancer then activate gene transcription through interaction with the transcription preinitiation complex bound at the promoter (composed of RNAPII and the mediator complex). Active enhancers are characterized by acetylation of H3K27, and in some cases trimethylation of H3K4, which causes the opening up of chromatin and accessibility of the DNA double helix to TF binding. *Figure made in Biorender and adapted from (Carullo & Day, 2019).*

1.1.4. Strategies to identify causal SNPs and determine underlying mechanisms in non-coding regions

Despite the challenges discussed, there have been sporadic successes (Alsheikh et al., 2022), and recently the field as a whole has seen some progress through the application of a more systematic methodology for tackling the problem (Claussnitzer et al., 2015; Sinnott-Armstrong et al., 2021). The approach, put forth by Claussnitzer et al. (2015), specifically addresses the mechanistic limitations of GWAS through a number of bioinformatic and experimental advances, a selection of which are described below (Figure 1.3).

To determine likely gene-regulatory regions that may be affected by a SNP within an LD-block, the surrounding locus can be assessed for chromatin status (Figure 1.3 A) (Carullo & Day, 2019; Spicuglia & Vanhille, 2012). Enhancers are regulated by epigenetic marks of the surrounding histone tails; active enhancers are characterized by a high density of H3K27 acetylation (H3K2ac) marks, which opens up the chromatin and makes the DNA accessible for TF-binding, while the same sites in inactive enhancers tend to be highly trimethylated and the chromatin compacted (Carullo & Day, 2019; Spicuglia & Vanhille, 2012). This is in contrast to promoters, which are recognized by a high density of H3K4me3 (Spicuglia & Vanhille, 2012). Thus, analyzing the epigenetic status of the region can give an indication of the type of regulatory element affected. Furthermore, since chromatin status varies between cell types, and regulatory regions are activated on a cell type-specific basis, assessing a wide variety of cell types (e.g. via a publicly available data such as that produced by the Roadmap epigenomics consortium (Roadmap Epigenomics et al., 2015)), affected cell type candidates can also be established (Figure 1.3 A) (Claussnitzer et al., 2015).

The candidate cell type can then be used as the host cell for *in vitro* luciferase reporter assays to narrow down the causal region of the LD-block: The block is divided into smaller segments (tiles), and each tile is cloned into a luciferase vector with a minimal promoter to detect tiles promote luciferase gene expression (Figure 1.3 B) (Claussnitzer et al., 2015). If one or more tiles gives positive results, the SNPs within the tile can be sequentially mutated to the alternative allele, and then reassessed by luciferase assay to determine whether the altered genotype of the SNP affects the enhancer activity of the tile. Based on such an assay, candidate causal SNPs are selected (Figure 1.3 C). These SNPs can be further assessed by analyzing the surrounding sequence for overlapping TF binding sites where one of the alleles disrupts the binding site. Any TF binding sites for which these criteria are met also suggest potential upstream mediators (Figure 1.3 C).

To further the search for causal SNPs (Figure 1.3 C), Claussnitzer et al. (2014) have developed a computational method to narrow down the number of potential candidates. The method is based on the fact that regulatory elements tend to have conserved TF binding site patterns (collectively termed cis-regulatory modules, or CRMs), such that, despite interspecies divergence of the exact TF recognition sequences, the pattern of sequential TF binding sites at a locus is conserved. The model, named phylogenetic module complexity analysis (PMCA), can be used to identify which SNPs are most likely to be causal by eliminating those that are outside of conserved CRMs. This can greatly reduce the number of SNPs necessary in need of experimental validation.

In search of potential target genes (Figure 1.3 D), expression quantitative loci (eQTL) data may offer some guidance (e.g. via the GTEx portal), as it links non-coding DNA loci to variation in gene expression (often tissue- or cell type-specific), thus suggesting a possible regulatory relationship between the two, though it may also be indicatory of a more distant relationship (GTEx Consortium, 2013). Chromatin capture is a method that can detect physical interactions between sequentially distant DNA segments, and when performed in the cell type of interest may provide evidence of possible targets (Claussnitzer et al., 2015).

Once good candidates for the causal SNP(s) have been established, together with suggestions for the affected cell type, regulatory function, and potential upstream mediators as well as possible target genes, the hypothesis needs to be evaluated *in cellulo*, and the consequential cellular phenotype determined (Figure 1.3 E). The ideal method of *in cellulo* evaluation is to genetically edit the candidate SNP from one allele to the other, so that two genetically identical cell lines are generated except with respect to the genotype of the SNP is question (Claussnitzer et al., 2015). Thus the cells can be compared for differences in gene expression (e.g. via mRNA sequencing or qPCR of suspected target genes), as well as assayed for detectable phenotype. In practice, *in situ* genome editing in mammalian/human cells has been a long-standing hindrance at this stage, with the major challenge being to achieve the specific edit in an efficient and reliable manner without causing off-target effects that may serve as confounders. However, recent developments in genome editing technology hold the potential to provide the catalyst the field has been looking for.

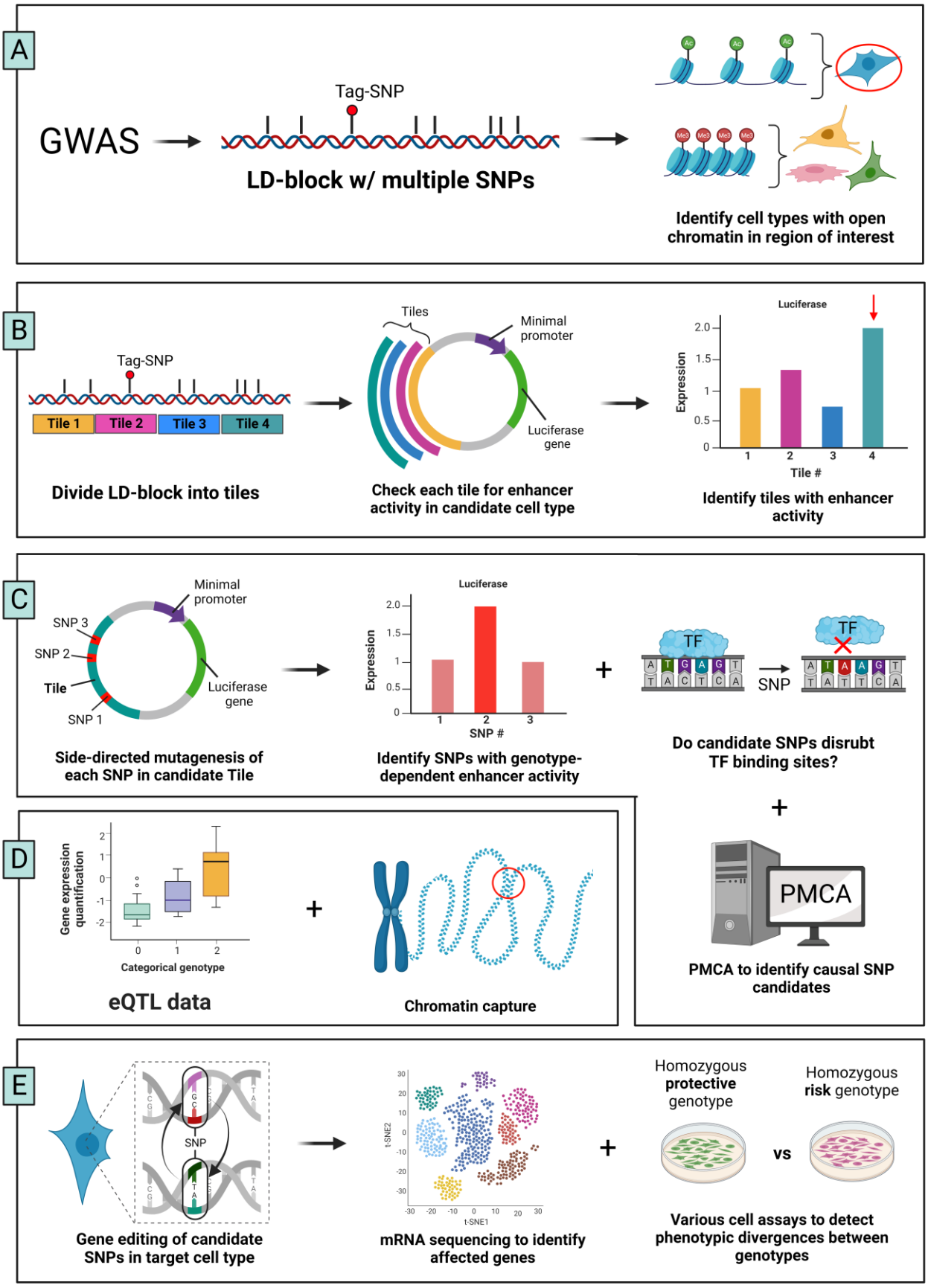

A
GWAS
Tag-SNP
LD-block w/ multiple SNPs
Identify cell types with open chromatin in region of interest
B
Tag-SNP
Tile 1
Tile 2
Tile 3
Tile 4
Divide LD-block into tiles
Tiles
Minimal promoter
Luciferase gene
Check each tile for enhancer activity in candidate cell type
Luciferase
Expression
Tile #
Identify tiles with enhancer activity
C
Minimal promoter
SNP 3
SNP 2
Tile
SNP 1
Luciferase gene
Side-directed mutagenesis of each SNP in candidate Tile
Luciferase
Expression
SNP #
Identify SNPs with genotype-dependent enhancer activity
TF
SNP
Do candidate SNPs disrubt TF binding sites?
D
Gene expression quantification
Categorical genotype
eQTL data
Chromatin capture
PMCA
PMCA to identify causal SNP candidates
E
SNP
Gene editing of candidate SNPs in target cell type
t-SNE2
t-SNE1
mRNA sequencing to identify affected genes
Homozygous protective genotype
vs
Homozygous risk genotype
Various cell assays to detect phenotypic divergences between genotypes

Figure 1.3 | Outline of how to systematically approach variant-to-function studies for non-coding DNA loci. A) Analyzing the surrounding chromatin structure of a GWAS-associated haploblock in various cell types can indicate regulatory function as well as cell types in which it is active. **B)** Do narrow down the causal region of the LD-block, it is divided into smaller segments (tiles) and each is assessed for enhancer activity in *in vitro* luciferase reporter assays (using the candidate target cell types as hosts). **C)** In tiles with detected enhancer activity, each SNP is sequentially mutated to the alternative allele and reassessed for genotype-dependent effects on enhancer activity. Identified candidate SNPs are further assessed in silico by analyzing the site for TF binding motifs that become ablated by the presence of the alternative SNP. Identified TFs are good candidates for upstream mediators of enhancer activation or repression. PMCA analysis can further narrow down causal SNP candidates. **D)** The analysis of readily available tissue-specific eQTL data can suggest affected downstream gene regulation. Chromatin capture studies in the affected cell type can be used to suggest direct downstream target genes. **E)** Once a good candidate system has been suggested, the hypothesis can be evaluated *in cellulo*, by genetically editing the suggested causal SNP to create two identical cell lines, except that one is homozygous for the risk allele and the other for the protective. Thus the two cell lines can be compared and contrasted for differential gene expression and phenotypic variation. *Method adapted from (Claussnitzer et al., 2015), and figure prepared in Biorender.*

1.1.5. CRISPR-Cas9

The state-of-the art gene editing tool is the engineered CRISPRS/Cas9 system, which consists of an RNA programmable endonuclease that can be targeted to almost any target DNA region (Babu et al., 2021). The system was first discovered in *Streptococcus pyogenes* as part of the bacterium's natural defense mechanism against pathogenic foreign DNA, and has since been developed as a tool for *in situ* DNA manipulation. The nuclease is complexed with a small RNA molecule containing a sequence complementary to its target, and thus guides the complex to the genomic target, at which a double-strand break is induced. The beauty of the CRISPR-Cas9 system lies in its programmability; the guide RNA (gRNA) can be designed to direct the nuclease to almost any 20 nt target DNA sequence (termed the "protospacer"), as long as it is directly upstream of a protospacer-adjacent motif (PAM); NGG (Figure 1.4).

Cas9 is a large protein with multiple functional domains, including a helicase, two nuclease-domains, a PAM interacting domain, and gRNA recognition domains (Babu et al., 2021). Structurally, the domains are organized into two lobes, between which a segment of the gRNA molecule is situated (Babu et al., 2021). This 20 nt sequence, termed the "spacer" (and identical to the protospacer target), is what guides Cas9 to its target DNA location, as it will only cleave DNA that is complementary to this sequence. To search the DNA for a complementary match, the gRNA-Cas9 complex locates and binds to PAM sequences distributed throughout the DNA, then unwinds the upstream DNA and positions the target strand into the groove adjacent to the spacer. If the target strand perfectly aligns with the RNA and a hybridized helix is formed, a conformational change occurs in Cas9, thus triggering DNA cleavage (Babu et al., 2021). The protein has two nuclease domains; RuvC which cleaves the protospacer sequence, and HNH which cleaves the complementary strand complexed with the sgRNA. In addition to the (variable) spacer sequence, the guide RNA consists of a scaffold sequence which forms a specific hairpin structure that facilitates its binding with Cas9 (Babu et al., 2021).

Several engineered versions of Cas9 have also been designed, to enhance or disable certain functions (Anzalone et al., 2020); Cas9-nickase (nCas9) has one of the two nuclease domains disabled, so the enzyme only nicks the target DNA rather than producing a double strand break. Catalytically dead Cas9 (dCas9) has both nucleases disabled, and thus functions to locate the enzyme to a specific site without cutting the DNA at all. Efficient Cas9 (eCas9) has been enhanced to have improve binding kinetics with DNA, thus being able to search more quickly for target sequences (Anzalone et al., 2020).

By designing the sgRNA to contain a spacer sequence specific to a desired target, the CRISPR-Cas9 system has typically been used to knock out target genes (Anzalone et al., 2020). Once a double-strand break has been made, the cell will repair the break by way of non-homologous end-joining (NHEJ), which often causes the insertion or deletion of nucleotides, thus mutating and disrupting gene function (Anzalone et al., 2020). However, this method is not directly suitable for DNA edits. An early solution was thus to combine the CRISPR/Cas9 system with a donor template containing the desired edit flanked by sequences homologous to the target site, in an attempt to induce the cell to repair the DSB by homology-directed repair (HDR), such that the desired edit becomes incorporated (Anzalone et al., 2020). However, this method is cumbersome and time-consuming, as HDR is successfully induced in only a small percentage of cells, and there is a high risk of imprecise edits or other induced mutations. However, by harnessing the site-directing capability of Cas9 and pairing it with other DNA-manipulating proteins, more sophisticated methods of gene editing are rapidly being developed.

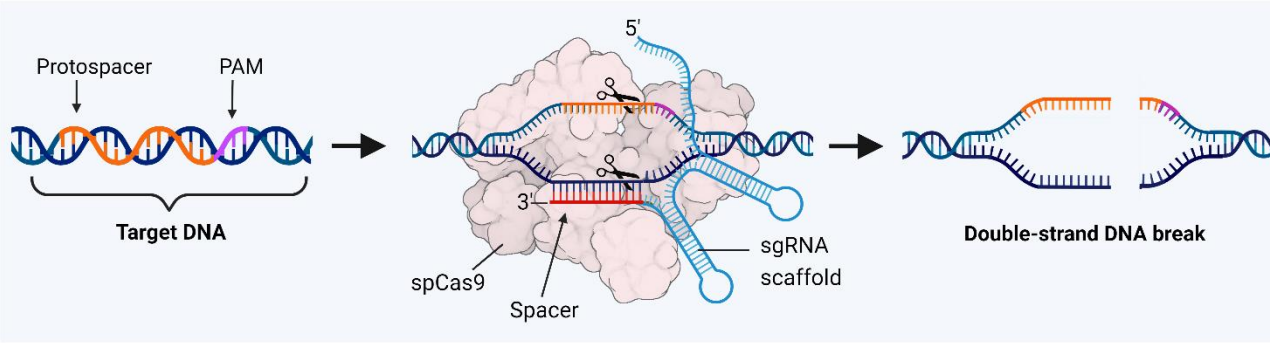


Figure 1.4 | The CRISPR-Cas9 system. The system consists of an RNA-programmable nuclease (SpCas9), complexed with the target-specific sgRNA. The sgRNA recognizes its genomic target on the basis of complementary pairing with the spacer sequence (red nucleotides of sgRNA), which thus is identical to the protospacer in the target DNA (orange nucleotides of target DNA). The sgRNA can be designed to target any 20 bp DNA target opposite a suitable protospacer. Protospacers must be directly upstream of a PAM-sequence (NGG; pink nucleotides of target DNA). Once the Cas9-sgRNA complex has located its target, a double-strand break is induced. *Figure produced in Biorender, adapted from (Anzalone et al., 2020).*

1.1.6. Base Editors and Prime Editing

In an attempt to leverage the CRISPR/Cas9 system to make more precise edits with higher success-rate, the base editors were developed, which are capable of inducing specific point mutation, by tethering a ssDNA deaminase enzyme to a catalytically deactivated Cas9 domain (Anzalone et al., 2020; Komor et al., 2016). The base editors (of which several varieties exists) are collectively capable of making all four base *transitions* (C ←→ T, and A ←→ G), which make up an estimated 30 % of human pathogenic variants (Anzalone et al., 2020). However, the system has several limitations when it comes to SNP targetability, and single-nucleotide transversions, as well as other mutation categories such as insertions, deletions and longer edits remained to be enabled (Anzalone et al., 2020).

Prime Editing (PE) is perhaps the most versatile genome editing tool developed to date, capable of precise *in situ* genome editing based on CRISPR-Cas9 technology (Anzalone et al., 2019). The PE system uses an engineered nCas9-reverse transcriptase (RT) fusion protein (the "Prime Editor") to induce a single-strand break in the protospacer-containing strand at the target site and reverse transcribe the desired edit in place of the original DNA sequence (Figure 1.5). To facilitate the edit, the prime editor utilizes a specially designed guide RNA; the pegRNA, which is an extended version of the sgRNA that includes an RT-template encoding the desired edit at its 5' end, preceded by a primer-binding sequence (PBS) that hybridizes with the sequence directly upstream of the nick induced by nCas9 (Figure 1.5, Step 3). Once the PBS has annealed to the nicked target strand, the exposed 3' hydroxyl group (OH) produced by the nick functions to prime the reverse transcription of the RT template by the RT domain (Figure 1.5, Step 4). Following reverse transcription, the prime editor dissociates from the target, and the nicked DNA strand contains a branched intermediate at the nicked site, with two redundant/overlapping flaps; the 3' flap containing the edit, and the 5' original sequence (Figure 1.5, Step 5). *In vivo* mechanisms must then remove one of the two flaps and repair the nick in the DNA. Although the 5' unedited flap will be more thermodynamically favorable to preserve since it is perfectly complementary to the reverse strand, multiple endogenous nucleases exist that selectively excise 5' flaps (e.g the endonuclease FEN1 and the exonuclease EXO1), and as such the edited flap has an increased chance of being preserved (Figure1.5, Step 6). In cases where the 5' flap is successfully excised/removed and the nick ligated (by DNA ligase), a mismatch will be present between the edited vs unedited strand. Endogenous DNA repair mechanisms will need to repair this, which means the edited strand has about 50 % chance of being preserved at this point (Figure 1.5, Step 7). This likelihood can be increased by introducing a nick on the unedited strand, using the PE-nickase function and a second, simple sgRNA (with no extension, referred to as the nicking guide RNA or ngRNA) that targets the edited strand and nicks the reverse strand. When this strategy is used, the system is referred to as PE3 (vs PE1 or PE2, depending on the generation of PE protein used). In some cases a suitable protospacer exists that overlaps with the edit-SNP, and thus the ngRNA can be designed to specifically target the successfully edited strand, thus preventing the nick from occurring too early (system referred to as PE3b) (Anzalone et al., 2019).

The PE method can edit single bases (both transitions and transversions) or multiple bases, as well as make insertions or deletions (Anzalone et al., 2019). However, the system does have a few limitations, including a requirement for the site to be edited to lie within editing range of a suitable protospacer (estimated to be ~10-35 bp, though in some cases longer distances may be possible) (Anzalone et al., 2019). Nevertheless, the wide range of edit categories recently made possible by these genome editors may prove to ameliorate the discussed challenged with in situ genome editing, and thus hold great potential to aid in the search of non-coding SNP function.

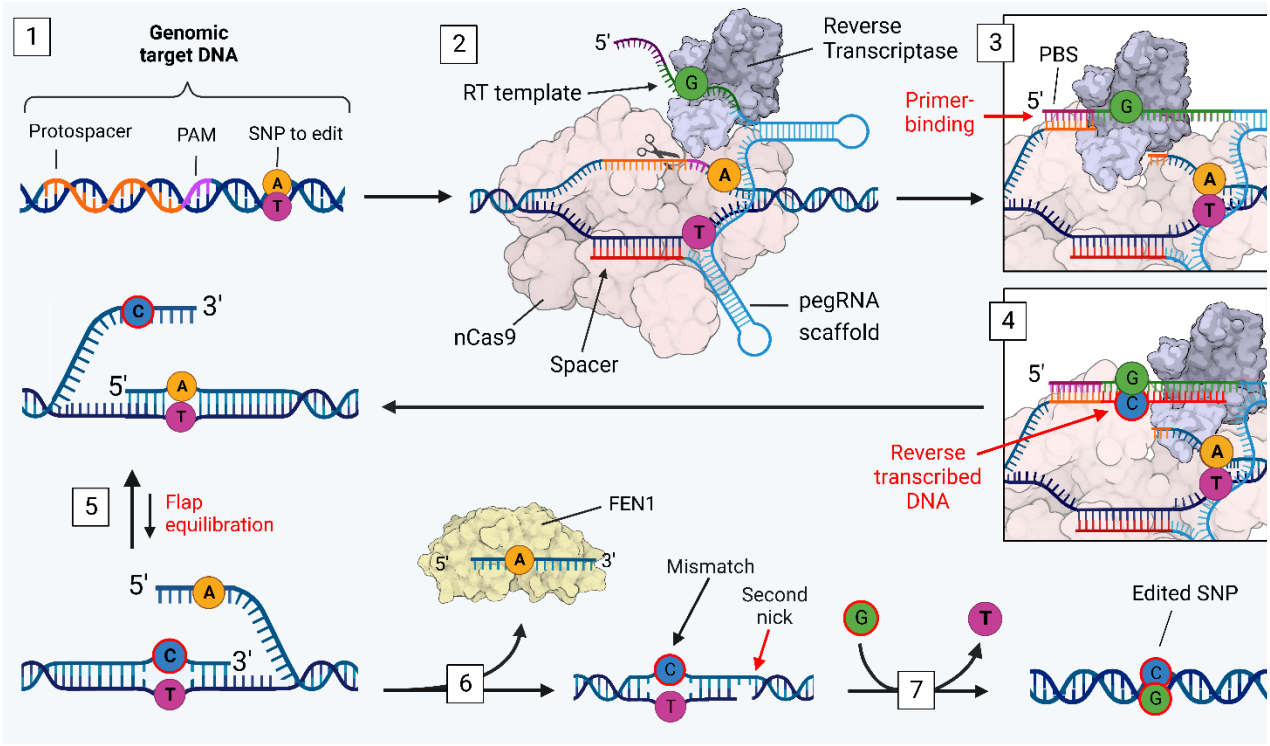


Figure 1.5 | Prime Editing mechanism. Step 1) PE is suitable for any genomic target (SNP) for which a unique PAM-adjacent protospacer sequence exists within editing distance (~30 nts). **Step 2)** The prime editor consists of an nCas9 subunit with a fused RT subunit at its C-terminal end. When complexed with the pegRNA containing a spacer sequence identical to the protospacer, the PE fusion protein is guided to the target region and hybridizes with the non-protospacer strand, nicking the other. **Step 3)** The pegRNA contains a primer-binding sequence (PBS) at its 5'-end which hybridizes with the nicked strand, and thus **4)** the free 3' OH of the nicked strand functions to prime reverse transcription of the RT template onto the target strand. **Step 5)** The prime editor dissociates, and the target DNA locus now contains two redundant flaps; the edited 3' flap and the non-edited 5' flap, which compete for hybridizing with the non-nicked strand. **Step 6)** Although the unedited flap is thermodynamically favored, it also has an increased chance of being excised by 5'excision nucleases such as FEN1. The nick is ligated by DNA ligase. **Step 7)** The resulting mismatch is repaired by endogenous DNA repair mechanisms, using the edited strand as template. Reprogramming the prime editor with a second sgRNA to nick the unedited strand increases the likelihood of the edit being conserved (prior to Step 7). *Figure made in Biorender and adapted from Anzalone et al (2020).*

1.1.7. CRISPR/Cas9-mediated epigenetic remodeling of target enhancer

In addition to all SNPs not being targetable by base editors or PE, another limitation in the application of these methods for GWAS-signal-to-function elucidations is that in many cases multiple SNPs in a predicted enhancer region may be suggested as potential causal SNPs, and thus targeting a single SNP may not be sufficient to show an effect, and making multiple edits increases the likelihood of off-target effects and the amount of work involved. To address these shortcomings specifically in enhancers, yet another method has recently been developed. The system, enhancer CRISPR-mediated activation/inactivation (CA/I), employs a dual effector system to modulate a target enhancer by epigenetically remodeling the chromatin structure, coupled with tethering transcriptional regulators to the site to promote or repress transcription (Figure 1.6) (Li et al., 2020). The system is similar to PE in that it also uses a Cas9-fusion protein to target a function of choice to a genomic target region. In this case, a catalytically deactivated Cas9 (dCas9) is fused with a chromatin remodeling domain; either the H3K27 acetyltransferase p300 for enhancer activation, or the H3K27 deacetylase LSD1 for inactivation (collectively referred to as effectors in this thesis). Further, a second effector (referred to as TFs for clarity) is associated with the system via the sgRNA, which has been engineered to include two MS2-hairpins; a structure which is recognized by the MCP RNA-binding proteins. In CA, a transcriptional activator VP64 domain is thus fused to an MCP domain, while for CI VP64 is exchanged with the transcriptional repressor KRAB (Li et al., 2020). The system can thus be used to both activate and repress the enhancer region in parallel populations of a target cell line, for potentially larger effect and more easily detectable change in downstream gene expression, though, of note, is that the results do not provide information about which of the SNPs are causal.

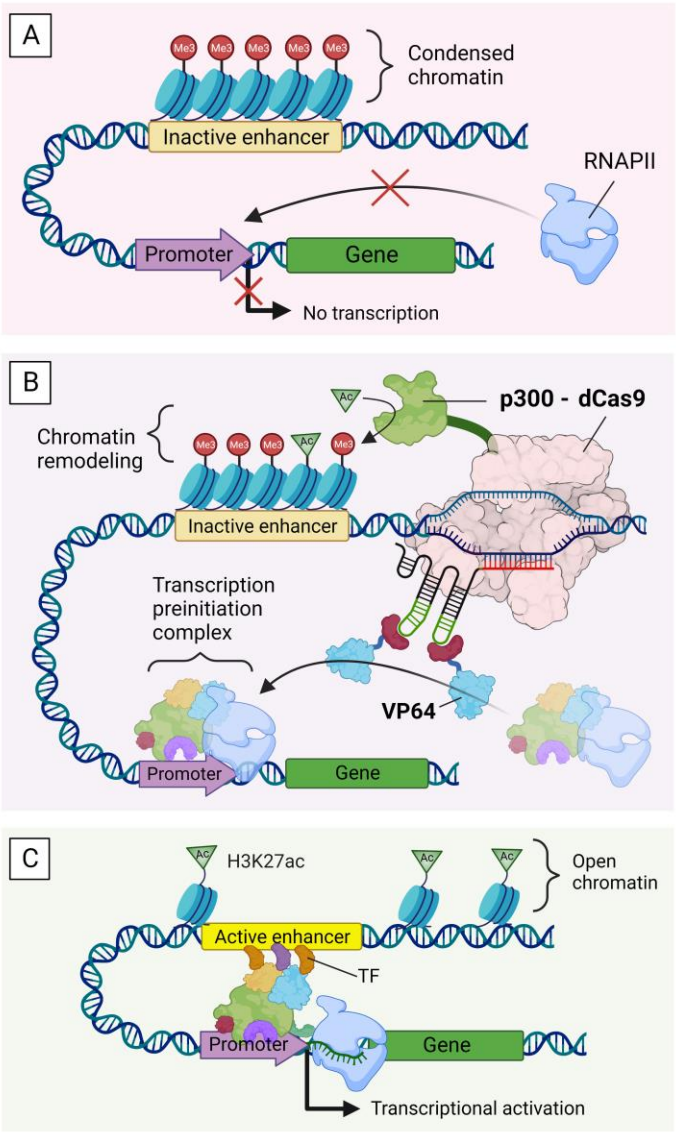


Figure 1.6 | Mechanism of CRISPR-mediated enhancer activation (CA). **A)** Enhancers are cell-type-dependently activated. Inactive enhancers are characterized by condensed chromatin that is regulated by histone marks (methylation of specific amino acids of the histone tails), and thus preventing TFs from accessing the enhancer. In enhancer-dependent gene transcription, the binding of TFs to the enhancer are necessary either to facilitate binding of the transcription preinitiation complex at the promoter, or for the activation of transcription from a poised transcription complex. **B)** In CA, dCas9 is fused with the histone acetyltransferase p300, that specifically acetylates H3K27, causing an opening up of the chromatin. When p300-dCas9 is guided to an inactive enhancer, the chromatin at the location becomes remodeled, and thus accessible to TFs. Further, the sgRNA has two associated VP64 subunits that initiate the assembly of the transcription complex at the promoter. **C)** The now activated enhancer can be bound by TFs, and together with the preinitiation complex activates transcription of down-stream genes. In CI, the system is reversed; dCas9 is fused with histone deacetylase LSD1, and sgRNA associates with two subunits of KRAB, which is a transcriptional repressor. *Figure made in Biorender and adapted from (Li et al., 2020).*

Collectively, the methods described above illustrate how to systematically approach the problem of converting GWAS signals to disease mechanisms, with the goal of elucidating the type of regulatory element affected, causal SNP(s), affected cell type, upstream regulators, and downstream targets (Claussnitzer et al., 2015). Furthermore, state-of-art CRISPR/Cas9-derived genome and epigenetic editing tools can be employed for *in situ* evaluation in the affected cell type for detection of downstream changes effects and characterization of the consequential phenotype and pathways involved.

1.2. Dissecting a novel GWAS signal associated with visceral obesity

To integrate prime editing and/or enhancer CRISPR activation/inhibition into the GWAS-to-function approach detailed above, our lab wished to apply this strategy to elucidate mechanisms behind genetic differences in susceptibility to obesity and obesity-related diseases. Obesity is for the most part a polygenic disease with a large number of co-morbidities, including type-2 diabetes (T2D), cardiovascular disease and cancer (WHO, 2015). It has received much scientific interest over the past several decades due to a tripling of its prevalence since 1975 (Collaboration, 2016; Iacobini et al., 2019; Ng et al., 2014), thus becoming a problem of global proportions affecting a skyrocketing 600 million people world-wide (WHO, 2015). Obesity is categorized on the basis of the body mass index (BMI), which functions as a proxy for body fat mass, and, as BMI increases above the healthy range of 20-25 kg/m^2, the risk of developing associated diseases rises drastically (Adams et al., 2006; Di Angelantonio et al., 2017), . In response to these worrying developments, some of the very first GWAS studies were aimed at elucidating genes associated with obesity (Frayling et al., 2007) and associated diseases (e.g. Type 2 diabetes [T2D] (Scott et al., 2007; Sladek et al., 2007)). There is, however, more to the story than simply BMI or total body fat percentage, as, on an individual level, there is a large variability in the degree of adiposity the body can sustain before metabolic consequences arise (Iacobini et al., 2019). For instance, BMI, though suitable for assessing populations, is a very crude measure when applied on an individual level, as it is simply a relative measure of weight to height and does not take other factors into account, such as muscle mass or fat distribution. Efforts have therefore been made to find more predictive markers for individual obesity-related health risk. This has led to the discovery that all body fat is not created equal, which, depending on where in the body it is stored, can have quite different health consequences (Kwok et al., 2016).

The main fat-storage organ of the body is the white adipose tissue (WAT), which is distributed in various depots across the body (Kwok et al., 2016). Depending on location, these depots can be categorized as either subcutaneous or visceral; the subcutaneous adipose tissue (SAT) is located directly under the skin, while visceral adipose tissue (VAT) is found within the abdominal cavity. The different fat depots display highly variable metabolic and secretory profiles, and thus contribute unequally to disease risk (Kwok et al., 2016). SAT is generally considered the biologically "intended" location for body fat stored as energy reserve, and, when within a healthy range, may be protective of disease. VAT, in contrast, functions to pad and protect the internal organs, and increased VAT mass has a more inflammatory secretory profile and is much more detrimental to health. Therefore, assessing body fat distribution and, more specifically, the volume of visceral fat mass, has been shown to be a more accurate predictor of individual health risk (Kwok et al., 2016).

Individual differences in fat deposition between the SAT and VAT is largely hereditary, but the genes and mechanisms involved are largely unknown (Sun et al., 2021), and therefore a prime target for a GWAS study. However, performing a GWAS screen on visceral obesity has been challenging, due to VAT mass being located internally and thus particularly difficult to measure accurately (Karlsson et al., 2019; Kwok et

al., 2016). The gold standards are assessment by CT-, MRI-, and DXA scans, but these methods are not very accessible for large numbers of individuals due to being costly, time-consuming, and requiring specialized personnel to perform (Karlsson et al., 2019; Kwok et al., 2016).

1.2.1. GWAS on predicted VAT mass

In 2019, the results of a GWAS on VAT mass were published where the authors had found a way around the difficulties of population-scale VAT mass measurements (Karlsson et al., 2019). Their solution was to develop a non-linear computational model to *predict* VAT mass based on easily accessible data, including height, weight, age, menopausal status in women, body circumference measurements (e.g. waist, hips, neck), and body composition measurements by bioimpedance. The model was first trained on 4 000 individuals for which the predicted VAT mass was verified with DXA scans, then the actual study was performed on an additional 400 000 individuals whose data was stored in the UK Biobank (UKBB).

The GWAS identified over 200 loci associated with increased predicted VAT mass, of which 102 were previously not linked to any adiposity-related phenotypes (Karlsson et al., 2019). Many of these loci were in LD-blocks that did not overlap with protein-coding regions, and thus the authors performed follow-up analyses to determine whether they may have a gene regulatory effect. The loci were analyzed for overlap with previously detected enhancer or promoter regions based on histone mark, DNase I hypersensitive sites, and TF binding site densities. Further, the group used GTEx eQTL data to determine which of the identified loci were also associated with altered gene expression in adipose tissue.

Based on this data, one of the GWAS-identified loci (11q23.3) was found to be associated with a genotype-dependent effect on expression of the gene HMBS (Karlsson et al., 2019), which has earlier been shown to have increased expression during adipogenesis/adipocyte differentiation (Moreno-Navarrete et al., 2017). The group therefore proposed that one or more of the SNPs in this region (and that were in LD with the tag-SNP identified in the GWAS), may be affecting the transcription of HMBS (Karlsson et al., 2019). To this end, the region was tested for genotype-dependent differences in promoter activity via a luciferase assay, through which they identified a locus containing a single SNP, rs1799993, suggesting it to be the causal SNP. Of the two alleles, the minor (A) allele showed a modest increase in luciferase expression relative to the major (C)-allele. This finding corresponded with the eQTL data, where the (A) allele is associated with higher HMBS expression relative to the (C)-allele. The authors of the study therefor hypothesized that the GWAS signal detected from the HMBS-locus was due to a genotype-dependent difference in promoter activity; that the adiposity risk allele (A) increased expression of HMBS, which further was driving visceral adipogenesis (Karlsson et al., 2019).

Although intriguing, this hypothesis has a major weakness: The promoter region examined in the study, which harbors rs1799993, transcribes a truncated isoform of HMBS which is only expressed in whole blood (Chretien et al., 1988), and thus this is likely not the causal mechanism for its association with increased

VAT mass. HMBS transcribes an enzyme involved in the biosynthesis of heme, an iron-binding cofactor of many enzymes and proteins, but the evidence of its involvement in any obesity-related phenotype is limited and contradictory (Moreno-Navarrete et al., 2017).

1.2.2. Identification of candidate causal SNPs in the visceral obesity-associated locus 11q23.3

Considering the lack of compelling evidence linking HMBS expression with adiposity, and the fact that the detected promoter is not likely active in adipose tissue, HMBS may very well not be the target gene of the regulatory region. Moreover, the LD-block of the VAT mass-associated tag-SNP is 30 kb and contains 38 SNPs in LD, and thus any these may be the causal SNP (Figure 1.7). Our lab therefore approached the question by adapting the approach presented by Claussnitzer et al. (2015), and thus hypothesized that the causal mechanism of the VAT mass-association may be that the locus functions as an enhancer in other tissues more relevant for adiposity than whole blood. Indeed, data from the Roadmap Epigenetic Consortium, which has characterized epigenetic data and chromatin structure of the human genome in 111 cell types and tissues (Roadmap Epigenomics et al., 2015), revealed that several regions of the LD-block has H3K27 acetylation marks characteristic of enhancers in several adipose tissue-related cell types, with the largest predicted enhancer being active in adipose-derived mesenchymal stem cells (AD-MSCs), but not in adipocytes differentiated from these cells (Figure 1.7).

The entire LD-block was then divided into tiles (segments of 3-4000 bp, Figure 1.7), and each tile was cloned into a luciferase reporter vector containing a minimal promoter and tested for genotype-dependent enhancer-activity (Mirza, 2022; Samuelsen, 2021). Using this approach, the 30 kb haploblock containing 38 SNPs in LD was reduced to a 1.8 kb tile (Tile 8a) containing only 5 SNPs, which overlapped with the annotated enhancer in adipose-derived MSCs (Mirza, 2022). The contribution of each of these SNPs was analyzed by site-directed mutagenesis of the reporter plasmid *in vitro,* followed by luciferase analysis. Strikingly, each of the 5 SNPs individually contributed to an increase in enhancer activity when altered from the VAT mass-associated protective variant to the risk allele, indicating that all 5 are potential causal SNPs (Mirza, 2022). PMCA analysis at this point can aid in ranking their relative contribution, and will be performed by Claussnitzer's lab. Meanwhile, each SNP was investigated for overlaps with TF binding sites, and rs1799993 has been found to be situated in the binding site of RARB, a known repressor of adipogenesis (Mirza, 2022; Schwarz et al., 1997). Intriguingly, the A (risk) variant of rs1799993 was predicted by JASPAR analyses to disrupt the binding affinity of RARB, thus attenuating the binding of the repressor to the enhancer, and, consequently, suggesting a potential explanation for the increased enhancer activity observed with the risk haplotype (Mirza, 2022). Due to rs1799993 being linked to this possible upstream mechanism, the SNP has been selected for downstream analyses.

Having established that AD-MSCs are the likely affected cell type, as well as having narrowed down the candidate SNP and determined the region that shows the strongest genotype-dependent enhancer activity,

the next step in elucidating the mechanism behind the association between this genomic region with visceral obesity is to study the effect of the SNP and the enhancer *in cellulo*.

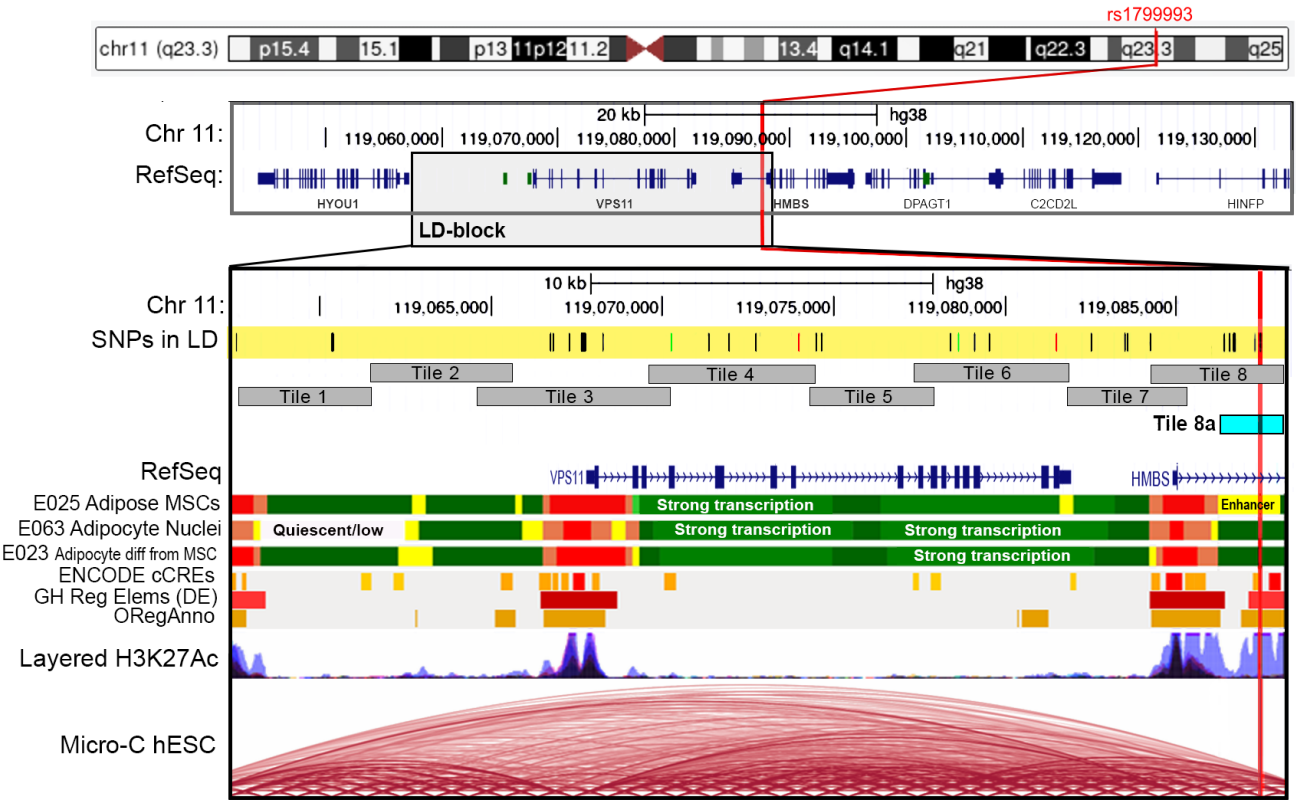


Figure 1.7 | Genomic location and analyses of the VAT mass-associated locus 11q23.3. Top: location of SNP rs1799993 is located on chromosome 11q23.3, as indicated by the red line. Middle panel: zoomed out view of the LD-block, including encompassed and adjacent genes. LD-block is by light grey box. Bottom panel: Zoomed in view of the LD-block. Chr11: Chromosomal locations. SNPs in LD: All LD-SNPs in the haploblock are indicated by vertical lines. SNP rs1799993 is highlighted in red. Below are grey rectangles indicating the approximate size and location of each of the 8 Tile-segments assessed for enhancer activity in a in vitro reporter assay (as depicted in Figure 1.3 B). Tile 8 gave the highest indication of enhancer activity by the assay, and was further dissected to produce Tile 8a (turquoise rectangle), which encompassed five LD-SNPs, all with genotypic affect on enhancer activity (see method of Figure 1.3 C). Tile8a was found to coincide perfectly with a predicted enhancer region in adipose-derived mesenchymal stem cells (E025 Adipose MSCs), and these were thus suggested as the affected cell type. Additional tracks shown in the panel: RefSeq – reference sequences of genes. E025 adipose MSCs, E063 Adipocyte Nuclei, and E023 Adipocyte differentiated from MSCs – epigenetic data of the region in the respective cell types, annotated based on the Roadmaps Epigenomics Project (yellow = enhancer regions, green = strong transcription, dark green = weak transcription, red = active transcription start-site). ENCODE cCREs: regulatory regions based on the ENCODE project (orange = enhancer-like signatures, red = promoter-like signatures). GH Reg Elems (DE): Dark red = predicted promoters (high confidence), light red = predicted promoters (lower confidence). ORegAnno: Orange = TF binding site. Layered H3K27Ac = compiled data for H3K27ac data from multiple cell lines. Darker color = region is acetylated in higher number of cell types. Micro-C hESC – indicates Linkage Disequilibrium. *Figure adapted from the master's theses of (Mirza, 2022; Samuelsen, 2021).*

Aims

This thesis is part of a larger project which aims to elucidate the causal SNP associated with visceral obesity, as well as identify downstream target genes. A single SNP, rs1799993, has been identified as a plausible candidate, via *in silico* and *in vitro* methods. In addition, epigenetic data suggest that the affected cell type is adipose-derived mesenchymal stem cells (AD-MSCs).

The current thesis aims to establish genome editing tools for manipulating the candidate causal SNP and/or the surrounding enhancer region in this cell type to enable subsequent identification of target gene(s) and downstream biological pathways.

Specific objectives include:

1. Designing sgRNA spacer sequences to target Cas9 to the locus for use with Prime Editing and enhancer CRISPR activation/inhibition and testing these spacers *in vitro* on the target DNA.
2. Genotyping rs1799993 in the target cell line, ASC52telo.
3. Preparing plasmids for *in cellulo* prime editing and CRISPR/Cas9-mediated enhancer activation/inhibition.
4. Optimizing transfection for the delivery of PE-plasmids to MSCs and perform a pilot editing experiment.
5. Prepare lentivirus for CRISPR/Cas9-mediated enhance activation/inhibition and test transduction of ASC52telo cells.
6. Establishing inducible p300-dCas9 and LSD1-dCas9 adipose-derived MCS cell lines for CRISPR/Cas9-mediated enhance activation/inhibition.

2. Materials

Table 2.1. | Online resources

Resource	Purpose	URL
Beacon Designer (Premier Biosoft)	Evaluation of primer-dimer and hairpin formation in PCR primers	http://www.premierbiosoft.com/qOligo/Oligo.jsp?PID=1
Biorender	Create scientific illustrations	www.biorender.com
BLAT (UCSC)	Sequence search in UCSC genome browser	https://genome-euro.ucsc.edu/cgi-bin/hgBlat?hgsid=289825942_OoRmdm12OuxgsoN1TO9Pmx0OWGq8&command=start
CRISPOR, v.4.99	Predict and evaluate sgRNAs (accessed via UCSC genome browser)	http://crispor.org
DeepHF	Predict and evaluate sgRNAs	http://www.deephf.com/#/cas9
NCBI Primer-BLAST	Primer design	https://www.ncbi.nlm.nih.gov/tools/primer-blast/
OligoEvaluator	Evaulation of primer-dimer and hairpin formation in PCR primers	http://www.oligoevaluator.com/
PrimeDesign	Tool used to design pegRNA component sequences for Prime Editing	https://primedesign.pinellolab.partners.org/
UCSC Genome Browser	Genome browser	https://genome.ucsc.edu/

Table 2.2. | Analytical software

Product	Supplier	Application/description
Fiji ImageJ	Open source	Image processing
FlowJo v10.9	BD Biosciences	Analyzing flow cytometry data
GraphPad Prism 9	Dotmatics	Statistical analysis
Image Lab TM Software V3.0	BioRad Laboratories, Inc	Gel visualization and annotation
IncuCyte 2021C	Sartorius	Live-cell imaging analysis
Microsoft Excel 2022	Microsoft	Data-handling
Microsoft Word 2022	Microsoft	Text editor
Adobe Photoshop	Adobe Creative Cloud	Graphics editor
SnapGene Viewer	Dotmatics	DNA sequence visualization and annotation
EndNote v 20.5	Clarivate	Reference management program

Table 2.3. | Instruments

Product	Supplier	Application
4D-Nucleofector System	Lonza/Amaxa	Nucleofection transfection
Gene Pulser Xcell Electroporation Systems	BioRad	Electroporation
IncuCyte S3 Live-Cell Analysis System	Sartorius	Live cell monitoring / fluorescence
Nikon Eclipse TS100 Microscope	Nikon	Fluorescence microscope/ cell imaging
SONY SH800 Cell Sorter	SONY	Flow cytometry/FACS
NanoDrop® ND-1000 Spectrophotometer	Thermo Scientific	Quantification of nucleic acids in solution
QIAexpert	QIAGEN	Quantification of nucleic acids in solution
GelDoc EZ Imager	BioRad	Agarose gel imaging

2.4. Oligonucleotides

Table 2.4.1. | Spacer sequences for sgRNAs

Spacer*	System	Sequence (5' → 3')	PAM	Strand	GC %	Distance** to SNP (nts)
Spacer-01	PE and CA/I	GAAACCAGAGGCAGCACTCT	AGG	(+)	55	9
Spacer-02a†	PE	TTCCATGTTGGTGAAACCAG	AGG	(+)	45	21
Spacer-02b†	CA/I	**G**TCCATGTTGGTGAAACCAG			50	
Spacer-04a	CA/I	ACTGCATGCATTAAGGACTG	AGG	(-)	45	173
Spacer-04b	-	**G**CTGCATGCATTAAGGACTG			50	
Spacer-06a	CA/I	CTGGTGGGACAGTGTACCCA	AGG	(+)	60	115
Spacer-06b	-	**G**TGGTGGGACAGTGTACCCA			60	
Spacer 10a	-	CTAGTCTTATGGAACATCTG	AGG	(+)	40	42
Spacer-10b	CA/I	**G**TAGTCTTATGGAACATCTG			40	
Spacer-ng(C)	PE3b	GT**t**CACCTAGAGTGCTGCCTC	TGG	(-)	57	15
Spacer-ng(A)	PE3b	GT**g**CACCTAGAGTGCTGCCTC	TGG	(-)	62	15

*names in **bold** indicate spacers selected for *in cellulo* use with indicated system.
†For Spacer2, version A was chosen for PE, while version B was chosen for CA/I.
**from cleavage site.

Table 2.4.2. | *Extensions for pegRNAs*

Extension	**Sequence (5' – 3')***	**Edits from → to**	**RT template length (nt)**	*PBS length (nt)*	*GC-content of PBS (%)*
Ext1_A	**TTCGTtCACCTAGA***GTGCTGCCTC TGGTT*	C→A	14	15	60.0
Ext1_C	**TTCGTgCACCTAGA***GTGCTGCCTC TGGTT*	A→C	14	15	60.0
Ext2_A	**TAAAGTTCGTtCACCTAGA***GTGCT GCCTC*	C→A	19	10	70
Ext2_C	**TAAAGTTCGTgCACCTAGA***GTGCT GCCTC*	A→C	19	10	70
Ext3_A	**TTCGTtCACCTAGAGTGCTGCCT CTG***GTTTCACCAACA*	C→A	26	12	41.7
Ext3_C	**TTCGTgCACCTAGAGTGCTGCCT CTG***GTTTCACCAACA*	A→C	26	12	41.7

***RT template in **bold** with SNP rs1799993 in red lowercase, PBS in *cursive*.

Table 2.4.3. | *PCR and sequencing primers*

Name/ID	**Sequence (5' → 3')**	**Binds to**	**Application**
641F	CAGGGTTATTGTCTCATGAGCGG	AmpR promoter	Sequencing, plasmid BPK1520 inserts
642F	GACTATCATATGCTTACCGT	pU6 promoter	Sequencing inserts of plasmids pegRNA-GG-acc, sgRNA_KRAB, and sgRNA_VP64
643F	TGTCTGATGTAGAAAGAAATGGGA	HMBS locus, forward	PCR of genomic DNA/ Genotyping rs1799993 in ASC52telo
644R	CTGGGGAACCTGTGCTGAG	HMBS locus, reverse	PCR of genomic DNA/ Genotyping rs1799993 in ASC52telo

Table 2.5. | Buffers and solutions

Stock	Composition	Application
Annealing Buffer	10 mM Tris-HCl (pH 8.5) 50 mM NaCl	Annealing complementary oligonucleotides
1X PBS	0.2 g/L KCl 0.24 g/L KH2PO4 1.44 g/L Na2HPO4 8 g/L NaCl pH adjusted to 7.4 with HCl	Cell culture
FACS buffer	1X PBS 2-5 % FBS 2 mM EDTA 2 mM NaN3	Flow cytometry
1X TAE buffer	40 mM Tris 20 mM Acetic acid 1 mM EDTA	Agarose gel electrophoresis
LB-medium	2.5 % (w/v) BD Difco Dehydrated Culture Media Adjusted to pH 7.5 by the addition of NaOH	Bacterial culture

Table 2.6. | Components for bacterial transformation and plasmid purification

Component	Supplier	Cat. no.
One Shot™ TOP10 Chemically Competent *E. coli*	Invitrogen	C404010
XL1- Blue Supercompetent Cells	Agilent	200236
One Shot® OmniMAX™ 2 T1 Phage-Resistant Cells	Invitrogen	C854003
Bacto™ Agar	BD Biosciences	214010
Glycerol, 85%	Merck	104094
Ampicillin sodium salt	CalBioChem	171254
S.O.C medium	Invitrogen	C4040
Difco™ Dehydrated Culture Media: Luria-Bertani (LB) Broth, Miller	BD Biosciences	244620
HiSpeed Plasmid Maxi Kit	QIAGEN	12663
QIAprep Spin Miniprep Kit	QIAGEN	27104

Table 2.7. | Plasmids

Plasmid name*	Size (kb)	Supplier	Application
pHR_TRE3G-**p300-dCas9**-P2A-mCherry	15.7	Addgene (Cat# 138456)	CA/I
pLenti_sgRNA(MS2)**_MCP-VP64**-IRES-zsGreen1	12.0	Addgene (Cat# 138461)	CA/I
pHR_TRE3G-**LSD1-dCas9**-P2A-mCherry	16.4	Addgene (Cat# 138460)	CA/I
pLenti_sgRNA(MS2)_MCP-**KRAB**-IRES-zsGreen1	12.1	Addgene (Cat# 138462)	CA/I
pSFFV-**Tet-on-3G-BFP**	10.5	Bo Huang, forwarded by Jian Xu	CA/I
pCMV-PE2-P2A-**GFP**	10.5	Addgene (Cat# 132776)	PE + Tr.opt.
pU6-pegRNA-GG-acceptor-empty-RFP	3.0	Addgene (Cat# 132777)	PE + Tr.opt.
BPK1520-empty-(sgRNA)	2.3	Addgene (Cat# 65777)	PE
pDG461-Cas9n(D10A)-2A-**EGFP**	9.7	Addgene (Cat# 100902)	Tr.opt.
pCAG-eCas9-GFP-U6-gRNA	10.2	Addgene (Cat# 79145)	Tr.opt.
pmaxGFP	3.9	Lonza	Tr.opt.
pCMV5-empty	Unknown	Unknown	Tr.opt.
Gag-Pol (2nd generation lentiviral packaging plasmid)	Unknown	Kind gift from Pouda Panahandeh	Lentivirus
VSV-G (2nd generation lentiviral packaging plasmid)	Unknown	Kind gift from Pouda Panahandeh	Lentivirus

Short names for reference **in bold.*
Tr.opt. = Transfection optimization.

Table 2.8. | Components used with DNA methods

Component	Supplier	Cat. no.	Application
Guide-it Complete sgRNA Screening System	Takara Bio	632636	sgRNA in vitro screening
NEBridge® Golden Gate Assembly Kit (BsmBI-v2)	NEB	E1602S	Golden Gate assembly
NEBridge® Golden Gate Assembly Kit (BsaI-HFv2)	NEB	E1601S	Golden Gate assembly
QIAquick Gel Extraction kit	QIAGEN	28706	Purification of PCR product
Platinum™ SuperFi II Green PCR Master Mix	Invitrogen	12369010	PCR
BigDye™ Terminator v3.1 Cycle Sequencing Kit	Applied Biosystems™	4337455	DNA sequencing
SeaKem LE Agarose	Lonza	50004	Agarose gel electrophoresis
Gel Loading Dye, Purple (6X)	NEB	B7024S	Agarose gel electrophoresis
GelRed™ Nucleic Acid Stain	MilliporeSigma™	SCT123	Agarose gel electrophoresis
GeneRuler 1 kb Plus DNA Ladder	ThermoFisher	SM1331	1.5% agarose gels
GeneRuler 50 bp DNA Ladder	ThermoFisher	SM0371	2% agarose gels

2.9. Cell culture components

Table 2.9.1. | Cell lines

Cell line	Description	Ethical approval number	Supplier (Cat. no.)
ASC52telo	Human adipose derived-mesenchymal stem cells from subcutaneous white adipose tissue, hTERT immortalized. Designed for stem cell research. (Wolbank et al., 2009)	EK589/2011, EK200/2005	ATCC® (SCRC-4000)
A41WAT	Human adipose progenitor cells sourced from subcutaneous white adipose tissue (neck), immortalized.	BIDMC, 2009-P-000101	Gift from the Joslin Diabetes Center, Boston, MA, USA.
HEK293T	Human endothelial kidney cells, immortalized by the addition of sheared adenovirus 5 (Berk, 2005). Hypotriploid with unstable genome. (Lin et al., 2014)	-	Gift from Nils Henrik Halberg, University of Bergen, Norway.

Table 2.9.2. | Cell culture components

Component	Supplier	Cat. no.
DMEM High Glucose	EuroClone	ECM0728L
DMEM/F-12, GlutaMAX™ Supplement	Gibco™	10565018
Mesenchymal Stem Cell Basal Medium	ATCC®	PCS-500-030
Mesenchymal Stem Cell Growth Kit for Adipose and Umbilical-derived MSCs – Low serum	ATCC®	PCS-500-040
Fetal Bovine Serum (FBS)	Gibco™	10270-106
OptiMEM	Gibco™	31985062
Trypsin-EDTA (0.25 %), with Phenol Red	Gibco™	25200056
Accutase	MerckMillipore	SCR005
Fibroblast Growth Factor (FGF) from bovine pituitary	Sigma-Aldrich	F3133
Epidermal Growth Factor, human recombinant (hEGF)	Sigma-Aldrich	E9644
Gentamicin	Gibco™	15750060
Penicillin and Streptomycin (PEST)	Sigma-Aldrich	P0781
Amphotericin B	Sigma-Aldrich	A2942
Incucyte® Nuclight Rapid Red	Sartorius	4717
Dimethyl Sulfoxide (DMSO)	EMD Millipore	317275
Nunc™ Cell Culture Dishes, Ø 10 cm	Thermo Scientific	168381
Nunc™ MicroWell™ 96-Well, Nunclon Delta-Treated, Flat-Bottom Microplate	Thermo Scientific	167008
Nunc™ Cell-Culture Treated Multidishes, 6 well	Thermo Scientific	140675
Gene Pulser/MicroPulser Electroporation Cuvettes, 0.4 cm gap	BioRad	1652088

Table 2.10. | Components for genotyping ASC52telo

Component	Supplier	Cat. no.
RNase A	QIAgen	19101
Proteinase K	QIAgen	19131
Chelex 100 Resin	Bio-Rad	1421253
Illustra™ ExoProStar™ 1-STEP Kit	Cyvita Life Sciences	US77705
Platinum™ SuperFi II Green PCR Master Mix	Invitrogen	12369010

Table 2.11. | Transfection reagents

Reagent	Supplier	Cat. no.
TransIT-X2® Dynamic Delivery System	Mirus	MIR 6005
TransIT ® -LT1 Transfection Reagent	Mirus	MIR 2305
X-tremeGENE™ 9 DNA Transfection Reagent	Roche Diagnostics	6365787001
Lipofectamine™ 2000	Invitrogen	11668027
Lipofectamine™ Stem Transfection Reagent	Invitrogen	STEM00001
P1 Primary Cell 4D-Nucleofector™ X Kit	Lonza	V4XP-1032

Table 2.12. | Components for lentivirus transduction

Component	Supplier	Cat. no.	Application
Lenti-X™ Concentrator	Takara Bio	631232	Lentivirus concentrator
Lenti-X™ GoStix™ Plus	Takara Bio	631280	Lentivirus quantification
Polybrene Infection/ Transfection Reagent	MerckMillipore	TR-1003-G	Transduction reagent
Doxycycline	Sigma Aldrich	D5207	Induces Tet-on-3G System

Table 2.13. | Flow cytometry

Component	Supplier	Cat. no.	Application
LIVE/DEAD™ Fixable Green Dead Cell Stain Kit, 488 nm	Invitrogen	L23101	Dead cell marker

3. Methods

3.1. Ethics and approvals

Given that that the work of this thesis involves the use of biological material from human subjects, ethical considerations must be taken into account. The project has been approved by the Norwegian Regional Ethics Committee (REK), including the use of genomic DNA samples sourced from the Western Norway Obesity Biobank (WNOB), approved by REK 2010/502 and REK 2018/2020. All patients have given written informed consent. The development and commercialization of the ASC52telo immortalized cell line (ATCC SRC-400) for research purposes was approved by the Ethics Committee of the University of Vienna (EK589/2011) and Ethikkommission der Landes Oberösterreich (EK200/2005).

3.2. Primer and sgRNA design

3.2.1. Design of spacer sequences for SpCas9 guide RNAs

Both PE and CA/I rely on CRISPR/SpCas9, and consequently require a guide RNA (gRNA) to direct the SpCas9 (Cas9) nuclease to the desired genomic location. While the overall gRNA structure differs between CA/I and PE due to different scaffold sequences, both types of gRNAs contain a spacer sequence complementary to the genomic target sequence that determines the Cas9 binding specificity. Thus, the same basic principles were applied when designing the spacer sequences for PE and CA/I.

The spacer sequences were identical to the genomic protospacer sequence of the target region, which is a 20 nt sequence directly upstream of a SpCas9 PAM sequence (NGG). Optimal spacers displayed high target specificity (to avoid off-target effects), as well as high cleavage efficiency (Haeussler et al., 2016). The former is influenced by the uniqueness of the spacer sequence, as well as the GC-content (GC-contents above 75 % have been shown to produce a high number of off-target cleavages), and was assessed using the MIT specificity score (Haeussler et al., 2016). The cleavage efficiency is dependent on sgRNA expression, conformation, and stability, with specific criteria varying depending on whether the sgRNA is transcribed *in cellulo* (U6 promoter) or *in vitro* (T7 promoter) (Haeussler et al., 2016; Riesenberg et al., 2022). Both promoters prefer a guanine (G) for efficient transcription initiation, and, since the spacer is located at the 5' end of the sgRNA, starting the spacer sequence with a G is usually recommended (Mali et al., 2013; Ranganathan et al., 2014). Therefore, for all spacers that did not naturally start with a G, a modified version (version B) was additionally designed where the first nucleotide was changed to a G. The Doench and the Moreno-Mateos algorithms were used for assessing predicted cleavage efficiency *in cellulo* and *in vitro*, respectively (Haeussler et al., 2016).

Candidate spacers were identified via the "CRISPR-targets" track of the UCSC Genome Browser (Human GRCh38/hg38). The track displays all potential spacer sequences for a DNA target region and provides a visualization of the predicted efficiency based on the algorithms described above (scores provided

by the online tool CRISPOR.org (Haeussler et al., 2016)). Only guides with both high specificity and efficiency (green color) were considered. These candidate guides were next manually ranked according to a merged *in cellulo* and *in vitro* specificity score For CA/I, the top 10 highest-scoring guides within a 1500 bp window centered on the predicted enhancer in Tile 8a (containing rs1799993) were selected for further analyses. For PE, a spacer targeting rs1799993 (Spacer-01, Table 2.4.1, materials), which valued proximity to the SNP ahead of cleavage efficiency, had previously been manually designed in our lab (Samuelsen, 2021). In this thesis, the UCSC browser was exploited to search for additional spacers with high efficiencies, within the PE editing range (+/- 35 bp) of rs1799993. All candidate sequences based on the UCSC genome browser screen were additionally evaluated using a separate online sgRNA design tool; DeepHF, which predicts cleavage efficiency of target DNA and potential off-target effects according to a separate algorithm (Wang et al., 2019). Only guides with high scores using both methods were selected for functional testing.

3.2.2. pegRNA design

In addition to a spacer sequence, pegRNAs for use with Prime Editing (PE) must also contain a custom template encoding the desired edit (RT template), as well as a primer binding sequence (PBS), which binds to the 3' end of the nicked top strand (Figure3.1). Together, these two elements are referred to as the "extension". The following guidelines suggested by Anzalone et al. (2019) were considered when designing the extension: PBS had a GC content of 40-60 %, and were 12-16 nts long. The RT templates covered the edit site and were kept between the recommended 10-35 nts. Moreover, a cytosine in the first nucleotide from the 5' end of the extension was avoided to prevent pairing with scaffold nucleotide G81, which would disrupt the sgRNA scaffold structure (Anzalone et al., 2019). Manually designed pegRNAs were validated and/or edited using the online pegRNA design tool PrimeDesign (Hsu et al., 2021).

An extension for Spacer-01 had previously been manually designed in our lab (Ext_1, Table 2.4.2, materials) (Samuelsen, 2021). In this thesis, an alternative extension (Ext_2) with a shorter PBS and longer RT-template was designed for the same spacer. Moreover, a second spacer (Spacer-02) with a single extension design was also constructed in this thesis. Moreover, for each extension, two versions were made, for editing from risk to protective and from protective to risk variants of rs1799993 (Table 2.4.2, materials). PrimeDesign also suggested spacers for use with the PE3b system, and thus a spacer sequence (Spacer-ng) was designed to guide a nick of the unedited DNA strand. This spacer sequence overlapped with rs1799993, so that it would selectively nick the strand prior to being edited (Table 2.4.1, materials).

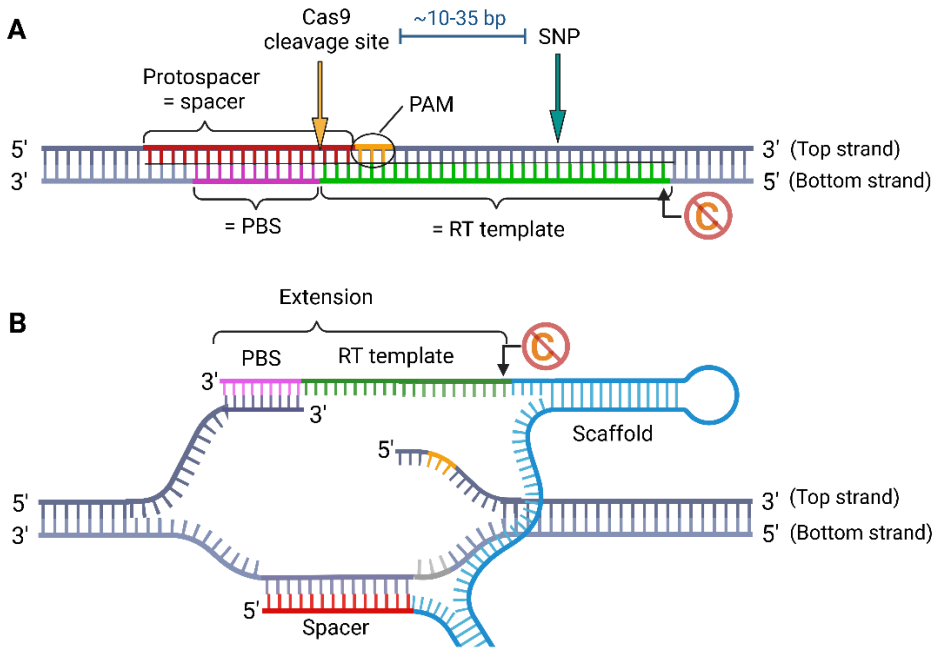


Figure 3.1 | pegRNA design. A) The various elements of the pegRNA were designed from the complementary sequences to where they would bind at the genomic target site. From top strand (also referred to as "forward strand", "cleavage target strand", or "edit strand"): The spacer sequence should be identical to the protospacer (red nucleotides), which is defined as the 20 bp sequence upstream of the PAM (NGG, yellow nucleotides), and thus binds to the complementary sequence on the reverse strand. The Cas9 cleavage site is located 3 bp upstream of the PAM sequence (yellow arrow, top strand only). For successful edit, the Cas9 nick should be within 10-35 bp of the edit site (SNP). Bottom strand (or reverse strand): PBS is identical to ~12-16 bp downstream of nick (on opposite strand), while RT template is identical to the sequence starting a few bases upstream of the SNP to be edited (but avoid starting with a cytosine) at the 5' end, and continuing downstream to end opposite the induced nick (3' end). **B)** Diagram showing how the pegRNA binds the target DNA following nicking of the top strand. Elements of the pegRNA are color coded as their identical sequences in A). (Number of nucleotides not representative). *Figure adapted from (Anzalone et al., 2019) and created in Biorender.*

3.2.3. PCR and sequencing primers

Primers for PCR amplification and sanger sequencing were designed via the NCBI Primer Blast online tool using default settings with the additional consideration that Maximum Template Mis-priming and Maximum Self and Pair Complementarity were kept as low as possible. Primers were checked for cross-homology against the human Refseq Representative Genomes database and/or the full sequence of the plasmid/vector harboring the target sequence, depending on intended use. Melting temperatures (T_m), as well as predictions for primer-dimers and secondary structure were independently verified using OligoEvaluator online tool by Sigma-Aldrich.

3.3. *In vitro* screening of gRNA cleavage efficiency

The cleavage efficiency of the selected sgRNA spacers was assessed *in vitro* using the Guide-it Complete sgRNA Screening System by Takara Bio. In brief, each spacer was transcribed in-frame with an sgRNA scaffold by T7 polymerase *in vitro*, then loaded onto recombinant purified SpCas9, and finally added to the target cleavage template generated by PCR amplification of genomic DNA for cleavage. The results were visualized and quantified via agarose gel electrophoresis of the cleavage reaction (Figure 3.2, and as detailed in the sections below).

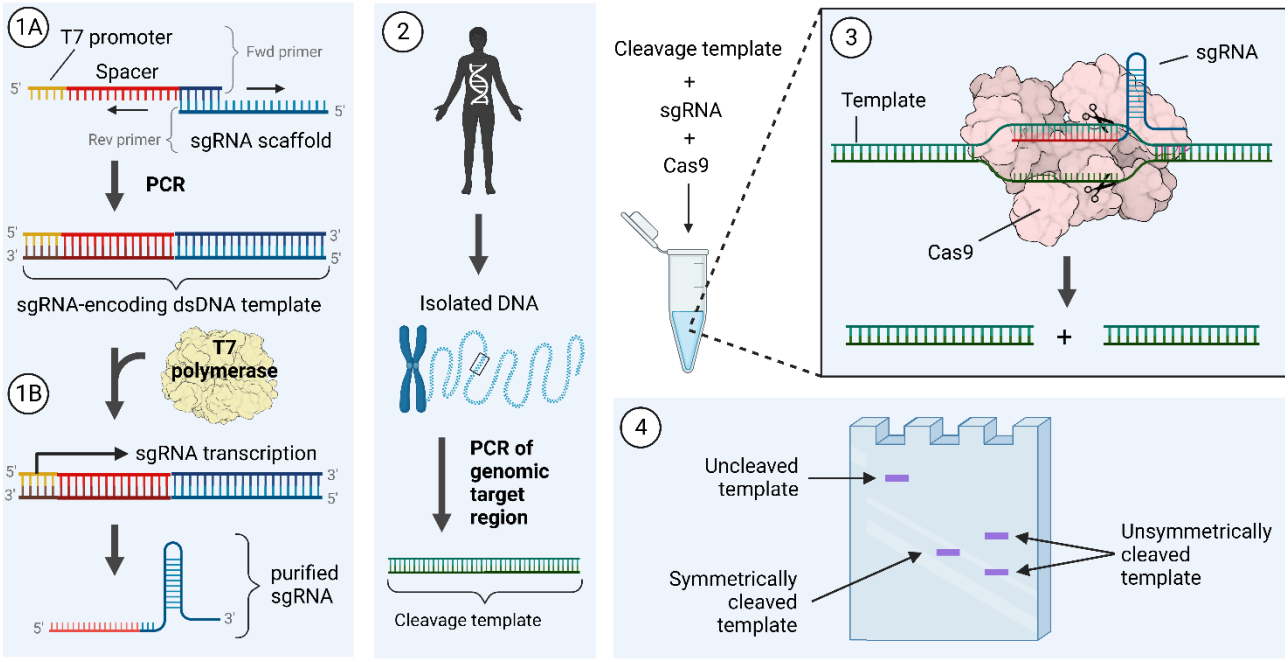


Figure 3.2 | Workflow for Guide-it *in vitro* Cas9 cleavage assay. Step 1A) sgRNA-encoding templates were generated from a PCR reaction, where the fwd primer consisted of (from 5' end): the T7 promoter (yellow), the target-specific spacer sequence (red), and a section of the sgRNA scaffold for annealing with the rev primer (blue). The rev primer consisted of the complementary scaffold sequence (light blue). **Step 1B)** sgRNAs were transcribed *in vitro* from the sgRNA-encoding templates by T7 polymerase, and purified. **Step 2)** A cleavage template was generated by PCR amplification of the genomic target region (~900 bp). **Step 3)** The purified sgRNAs were loaded onto recombinant, purified sqCas9, and added to the cleavage template to assess on-target cleavage efficiency. **Step 4)** The results were analyzed by agarose gel electrophoresis. The uncleaved template was expected to run higher in the gel than the two fragments generated by successful cleavage, which would generate a single band for spacers targeting the center of the cleavage template, or two bands if the sgRNA target was situated more toward one end. Cleavage efficiencies were calculated based on percentage of cleaved template vs uncleaved generated by each spacer. Figure made with Biorender.com.

3.3.1. Generating sgRNA templates

sgRNA-encoding DNA templates were generated from a PCR reaction from specially designed primers, as described in Figure 3.1, Step 1A. 1-2 G's (depending on whether the spacer started with a G or not) were

added directly downstream of the promoter sequence to enhancer transcription initiation. Fwd primers were ordered from Sigma Aldrich, while rev primers were supplied with the kit. The PCR reactions (25 μL) were prepared in 1X PrimeSTAR Max Premix with 0.2 μM Fwd primer and 0.2 ng/μL Rev primer. The reaction was run in a thermocycler with 33 cycles of 10 s denaturation at 98 ºC followed by 10 s annealing/elongation at 68 ºC. The PCR products were verified by agarose gel electrophoresis (2 % gel, as described in Section 3.4.2).

3.3.2. sgRNA transcription and purification

The sgRNA templates were transcribed *in vitro* by T7 polymerase (Figure 3.2, Step 1B), in a reaction containing 1:4 (v/v) sgRNA PCR template and 25 U/μL Guide-it T7 Polymerase Mix in 1X Guide-it Transcription buffer, which was incubated for 4 h at 37 ºC. Following transcription, the template DNA was degraded by incubating the solution for 15 min at 37 ºC with 10 U recombinant DNase. The transcribed sgRNAs were then purified via RNA purification spin columns and a series of wash and centrifugation steps according to kit's instructions. The final products were eluted in 20 μL RNase free water, and RNA concentrations were measured by NanoDrop.

3.3.3. PCR amplification of gDNA template

The target cleavage templates were generated by PCR (Figure 3.2, Step 2) from two previously purified human genomic DNA samples, one (V602) homozygous for the protective allele of rs1799993 (C/C) and the other (V586) homozygous for the risk allele (A/A). Primers were designed using NCBI Primer-BLAST to amplify a 891 bp fragment surrounding the SNP and covering the binding site of all spacers for both PE and CA/I, such that cleavage would result in fragments of either approximately equal size, or two unequal sizes (Figure3.2, Step 4). The PCR reactions were run in 50 μL 1X Terra PCR Direct Buffer, containing 15 pmol each of Fwd (642F) and Rev (643R) primer (Table 2.4.3), 1.25 U Terra PCR Direct Polymerase, and 30 ng genomic DNA. The reactions were run in a thermocycler with initial denaturation of 2 min at 98 ºC, followed by 35 cycles of 10 s denaturation at 98 ºC, 15 s annealing at 60 ºC, and 1 min elongation at 68 ºC, and final hold at 4 ºC. The products were verified via agarose gel electrophoresis (1.5 % gel, Section 3.4.2), then purified by gel extraction (Section 3.4.2).

3.3.4. SpCas9-sgRNA in vitro cleavage assay

The transcribed and purified sgRNAs were assayed in an *in vitro* reaction with recombinant SpCas9 nuclease and the PCR-amplified genomic DNA cleavage template (Figure 3.2, Step 3). For each reaction, 50 ng transcribed and purified sgRNA was loaded onto 250 ng Guide-it recombinant Cas9 nuclease by incubating for 5 min at 37 ºC (1.5 μL reaction). This was then added to a reaction containing 1X Cas9 Reaction Buffer containing 1X bovine serum albumin (BSA) and 100 ng gDNA template, which was incubated for 1 h at 37

ºC, then denatured at 80 ºC for 5 min. A control reaction was run with the supplied control sgRNA and matching cleavage template according to manufacturer's protocol. Reaction results were analyzed via agarose gel electrophoresis (1.5 % gel, Section 3.4.2). To quantify the cleavage efficiency of each sgRNA, the quantification tool of the BioRad GelDoc EZ Imager software was used: For each sample, the densities of all bands (successfully cleaved template fragments and unsuccessfully cleaved whole template) were measured, and the combined density of the two cleaved products was compared to the density of the unsuccessfully cleaved band to generate a value of cleavage efficiency (% of template successfully cleaved).

3.4. PCR and sequencing of DNA

3.4.1. Polymerase chain reaction

Polymerase chain reactions (PCR), when not part of a kit, were run in 1X SuperFi II Green MaMix (20 µL total reaction volume) containing 10 ng DNA template and 10 pmol each of Fwd and Rev primers. The reaction was run in a thermocycler for 30 s at 98 ºC (initial denaturation), followed by 30 cycles of 10 s denaturation (98 ºC), 10 s annealing (60 ºC), and 30 s elongation (72 ºC), and a final elongation of 5 min at 72 ºC. Controls were run without template.

3.4.2. Agarose gel electrophoresis

Agarose gels for the electrophoretic separation of DNA molecules by charge/mass ratio, were prepared in 1X TAE buffer with either 1.5 % or 2.0 % (w/v) SeaKem LE Agarose, and 1:10 000 (v/v) GelRed nucleic acid stain. Samples were prepared in 1X (final) Purple Gel Loading Dye, with 5:6 (v/v) DNA sample. Gels were submerged in 1X TAE buffer prior to sample loading, then run at 100-150 V until dye-front had migrated about ¾ the length of the gel. Gels were imaged with BioRad GelDoc EZ Imager, and analyzed with the accompanying software.

3.4.3. Gel purification of PCR products

The QIAquick gel extraction kit was used to purify PCR products to remove any unspecific amplification products, as well as PCR buffer and enzymes. To do so, the entire PCR sample was run on a 1.5 % (w/v) agarose gel, and the band of interest was excised under UV light (mainly 365 nm wavelength, 254 nm wavelength used very sparingly). The DNA was then extracted via a spin column with a silica membrane following a standard bind-wash-elute procedure. An Eppendorf MiniSpin Centrifuge was used for all spin steps, and the purified DNA was eluted in TE buffer and quantified via NanoDrop.

3.4.4. BigDye sequencing

All sequencing reactions were run using the BigDye Terminator v3.1 Cycle Sequencing Kit, which uses the Sanger sequencing principle. The BigDye sequencing reactions were prepared in 1X Sequencing Buffer (10 µL total reaction volume) with the addition of 10 % (v/v) DigDye v3.1 Ready Reaction Mix, 3.2 pmol primer (Fwd or Rev), and 100-200 ng DNA template (depending on size). For each template, Fwd and Rev primers were run separately, and two replicates were run for every primer-template combination, such that each sequence was verified four times. Template-free reaction were run as controls. The reactions were run in a thermocycler for 30 s initial denaturation (96 ºC), then 35 cycles of 10 s denaturation (96 ºC), 5 s annealing (50 ºC), and 4 min elongation (60 ºC). Following run completion, samples were delivered to the Core facility at the Department of Medical Genetics, Haukeland University Hospital, where they were sequenced by capillary electrophoresis. The returned results were subsequently viewed/analyzed with SnapGene.

3.5. Cloning and purification of Prime Editing and CA/I plasmids

After designing sgRNAs spacers and testing them *in vitro*, the best-performing spacers were selected for *in cellulo* genome editing (PE) and/or epigenetic editing (CA/I). To this end, the spacers were cloned into different overexpression plasmids for PE and CA/I, respectively, transformed into competent bacterial cells and subsequently purified as described in the following sections.

3.5.1. Golden Gate assemblies

The principle of Golden Gate (GG) cloning (Engler et al., 2008) was used to assemble the guide RNA inserts into their respective expression vectors (Figure 3.3). GG cloning takes advantage of Type IIS restriction enzymes to linearize the empty vectors. Since Type IIS enzymes cleave DNA at a site adjacent to their recognition sites, the linearized plasmid will have two non-complementary 5’ overhangs, which prevents vector self-ligation, as well as repeated cleavage after ligation with insert. The insert is generated with complementary 5’ overhangs, which further enables correct orientation of the insert.

The pegRNAs for PE were assembled from three segments; the spacer sequence (Table 2.4.1), the pegRNA scaffold sequence, and the extension sequence (Table 2.4.2). Each of the three segments were ordered from Sigma Aldrich as ssDNA oligonucleotides (top and bottom) that were annealed prior to sequential assembly and ligation with expression vector. The top and bottom strand of each segment were designed to contain different 5’ overhangs (Figure 3.3 A) complimentary to those of the vector and/or next segment such that only vectors containing all three inserts in correct order and orientation would ligate into a circular plasmid (Figure 3.3 B). Full sequences listed in supplementary information.

The pegRNA-encoding sequences were cloned into the **pU6-pegRNA-GG-acceptor** plasmid **(Anzalone et al., 2019)** by BsaI digestion (plasmid map, Figure 3.4 A). The plasmid has two recognition sites

for BsaI, one on either side of an mRFP1 dropout-cassette. This allowed for easy selection of successfully assembled plasmids, as bacteria harboring undigested plasmid appeared red, while bacteria transformed with correctly assembled insert appeared white.

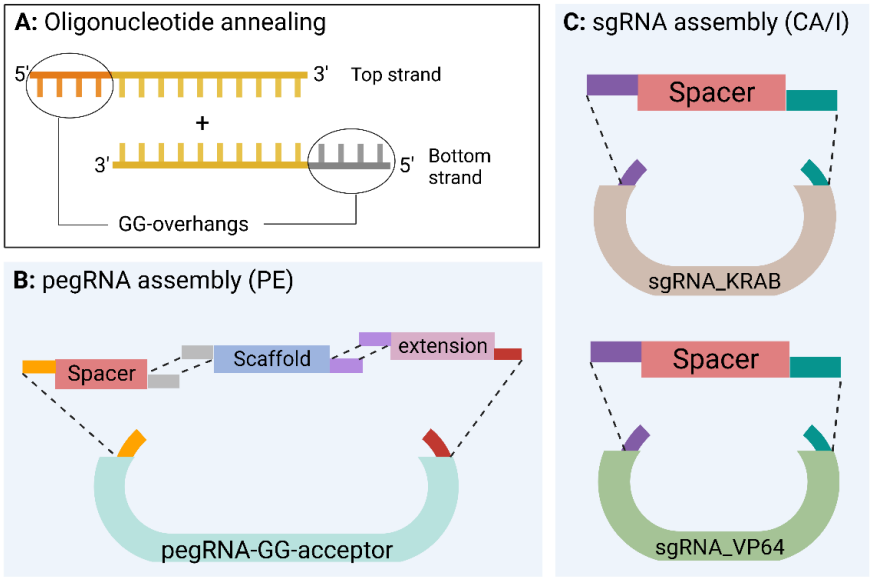


Figure 3.3 | Golden gate assembly of pegRNAs and sgRNAs. A) To generate inserts with 5' GG-overhangs, the top and bottom strands were ordered separately as ssDNA oligonucleotides, each with a unique 5' sequence, and subsequently annealed to generate dsDNA. **B)** pegRNA inserts were assembled from three dsDNA segments; spacer, scaffold, and extension. Each segment had specific 5' overhangs that enabled sequential assembly. The acceptor vector was linearized by a Type IIS RE, which generated two non-complimentary 5' overhangs (yellow and red). Thus only the correctly assembled plasmid with insert could be circularized and ligated. pegRNA-GG-acceptor, pU6-pegRNA-GG-acceptor-empty-RFP. **C)** Same principle was used for the Lentiviral sgRNA-plasmids, except that for these plasmids only the spacer sequence needed to be inserted. sgRNA_KRAB, Lenti-sgRNA(MS2)-MCP-**KRAB**-IRES-zsGreen1; sgRNA_VP64 and Lenti-sgRNA(MS2)-MCP-**VP64**-IRES-zsGreen1, respectively. *Figure made in Biorender.*

For CA/I, each spacer (Table 2.4.1) was cloned into two different lentiviral vector backbones (**Lenti-sgRNA(MS2)-MCP-*effector*-IRES-zsGreen1**, where *effector* is either an activating (VP64) or repressive (KRAB) coregulator (Li et al., 2020), Figure 3.4 C)). The plasmids have two adjacent recognition sites for the Type IIS restriction enzyme BsmBI, thus generating two non-complementary 5' overhangs when digested. Since the vectors already contained the engineered scaffold sequence with MS2 hairpins, only the spacer sequence needed to be inserted, and for this assembly the spacer sequences were designed with overhangs to match the digested vectors (Figure 3.3 C). Finally, the ngRNAs for the PE3b system were cloned into the **BPK1520-empty** vector (Figure 3.4 B) using BsmBI using the same principle as above.

Annealing reactions for complementary ssDNA oligos were prepared with 4 µM each of top and bottom sequence in annealing buffer (10 mM Tris-HCl (pH 8.5), 50 mM NaCl). The reactions were heated to 95 ºC for 3 min, then allowed to cool gradually to RT. A GG cloning reaction was then prepared for each

plasmid-insert combination, such that the vector was linearized and insert ligated in a single step: The reactions were prepared in 1X T4 DNA Ligase Buffer containing 3 ng/µL plasmid backbone, 0.1 µM annealed spacer (as well as 0.1 µM each of scaffold and extension for the PE plasmids), and 5 % (v/v) NEB Golden Gate Enzyme Mix (containing BsaI-HFv2 or BsmBI-v2, depending on target plasmid, as well as T4 DNA ligase). The GG cloning reactions were run in a thermocycler for 8 cycles alternating for 5 min each between 16 ºC (for T4 ligase action), and 37 ºC (for BsaI) *or* 42 ºC (for BsmBI), followed by an additional 15 min at 37 or 42 ºC and 15 min at 80 ºC for enzyme inactivation.

The assembled plasmids were subsequently transformed into competent bacteria (Section 3.5.2). For pU6-pegRNA-GG plasmids, the red/white selection marker was used to pick colonies with likely correctly assembled plasmids. These were grown to maxi-culture and purified, then sequenced via BigDye sequencing to verify that the assemblies were correct. The Lenti-plasmids had no screening marker, thus multiple transformed colonies were grown to mini-culture (5 ml), and 2 ml were purified by miniprep and sequenced with BigDye to check for insert. When plasmids harboring the correctly assembled plasmids were identified, the remaining 3 ml of mini-culture was amplified further to maxi-cultures and then purified by maxi-prep (see the following sections).

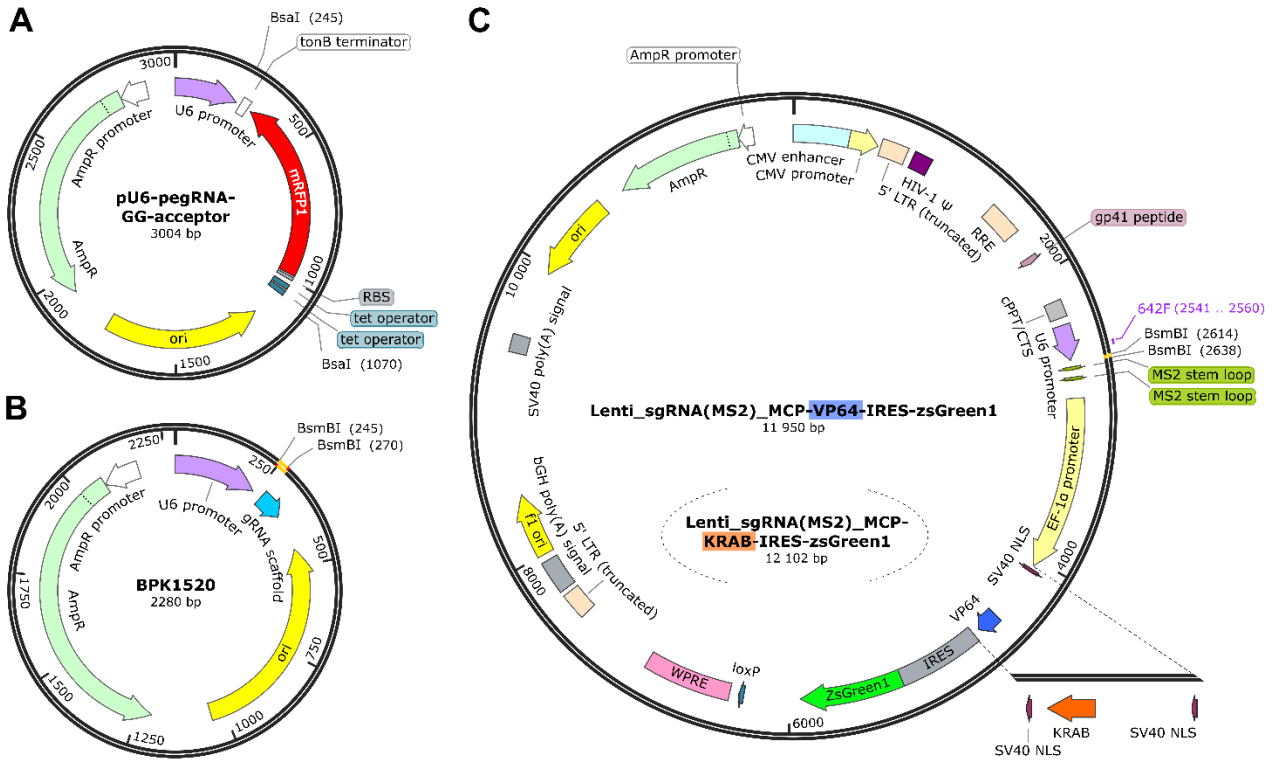


Figure 3.4 | Plasmids used with GG-assembly. A) pegRNA acceptor plasmid, digested with BsaI, which resulted in loss of the mRNF1 dropout cassette when successfully assembled. **B)** ngRNA acceptor plasmid, digested with BsmBI. **C)** Lenti-sgRNA plasmids, which were identical to one another except for that the one harbored the sequence for VP64 (dark blue) and the other KRAB (orange) in the downstream reading frame.

3.5.2. Bacterial transformation

Three strains of *Escherichia coli (E. coli)* competent bacteria were initially tested and compared for colony count and total plasmid yield following transformation and amplification: One Shot Top10, XL1 Blue, and One Shot OmniMax2 $T1^R$ cells. Since the OmniMax2 cells outperformed the other two strains, it was used for all subsequent transformations. Briefly, 25 µL competent *E.coli* cells were mixed with 0.5-1.0 µL plasmid DNA (in TE buffer) and incubated for 30 min on ice, followed by heat-shock at 42 ºC for 35 s, then returned to ice for 2 min. 125 µL SOC medium was added to each reaction, and the transformed bacteria were incubated for 1 h at 37 ºC whilst shaking at 225 rpm. Cells (30 µL and/or 60 µL) were then streaked on pre-heated (37 ºC) LB-agar plates with ampicillin (100 µg/mL) and incubated upside down overnight at 37 ºC. (All plasmids used in the present work carried the ampicillin resistance gene).

3.5.3. Amplification of transformed bacteria/cells

High-quality colonies (large, in an area of low colony density, and lacking satellite colonies) were picked for further cultivation. Mini-cultures were prepared by inoculating 2.5 ml LB-medium containing ampicillin with a single colony of transformed bacteria, and incubated for 6-16 h at 37 ºC while shaking (225 rpm). The mini-culture was then either purified via Mini-prep (Section 3.5.5) or used to inoculate a maxi-culture of 300 ml ampicillin-containing LB-medium and incubated for a further 16 h at 37 ºC, 225 rpm. Bacteria were harvested by centrifugation at 5600 x g for 20 min (4 ºC), and supernatant (SN) discarded. The pellet was then lysed and purified via Maxi-prep plasmid purification (Section 3.5.6). When necessary, pellets were stored at -20 ºC until use.

3.5.4. Glycerol stocks

Glycerol stocks of all transformed bacteria were maintained by freezing cultured transformed bacteria (from mini- or maxi-cultures) in 15 % (v/v) glycerol at -80 ºC.

3.5.5. Mini-Prep Plasmid Purification

Mini-preps were performed for use in sequencing reactions to determine bacterial cultures transformed with plasmids harboring the correct inserts. In these cases, the QIAprep Spin Miniprep Kit was used following the microcentrifuge protocol according to manufacturer's instructions. Purified mini-prep products were only used for sequencing and not transfections due to endotoxins not being removed by this method.

3.5.6. Maxi-Prep Plasmid Purification

The QIAGEN HiSpeed kit was used according to manufacturer's instructions to purify larger quantities of transfection grade, endotoxin-free plasmids from 300 mL maxi culture.

3.6. Cell culture

3.6.1. Cell lines

Three cell lines were utilized in the present thesis (Table 2.9). The main cell line, **ASC52telo**, is a human hTERT-immortalized adipose-derived mesenchymal stem cell line (AD-MSC), purchased from ATCC (SRC-4000). This cell line was the target for both PE and CA/I, and used in the majority of cell experiments. The cell line **A41WAT** was used early in the project as a proxy for ASC52telo in a couple transfection tests, before the latter had arrived, since A41WAT is a human adipose progenitor cell line with similar sensitivity to transfection as ASC52telo. The cell line **HEK293T** was used as a lentivirus packaging cell line, as well as a positive control cell line in transfection tests.

3.6.2. Cultivation and passaging

Cell lines were cultured in either 10 or 15 cm culture-grade dishes using cell line-specific medium (Table 3.1), and maintained in a humidified cell incubator at 37 ºC with 5 % CO_2. Growth medium was changed every 2-3 days, and cells were passaged when they reached 70-90 % confluence. To do so, the cells were harvested by trypsin (A41WAT and HEK293T) or Accutase (ASC52telo) and a subset of the cells were transferred to a new culture dish with the appropriate volume of growth medium. ASC52telo cells were always counted when passaged, and usually seeded at a density of 5000 cells/cm^2, while the other cell types were typically passaged at a 1:10 dilution. Cells were maintained for a maximum of 10 passages.

Table 3.1. | Composition of cell line specific growth medium

Cell line	Basal media	Supplements
ASC52telo	MSC Basal Medium *	MSC Growth Kit (2% (v/v) FBS, 5 ng/ml rhFGF basic, 5 ng/ml rhFGF acidic, and 5 ng/ml EGF final) Amphotericin B (0.25 µg/ml final) Gentamicin (100 µg/ml final)
A41WAT	DMEM/F-12, GlutaMAX™ supplement	FBS (10% v/v), FGF (0.1 µg/L) EGF (1.0 µg/L) Gentamicin (50 µg/ml)
HEK293T	DMEM High Glucose	FBS (10 % v/v) Penicillin (100 U/ml) Streptomycin (100 µg/ml)

3.6.3. Cell counting

The Invitrogen Countess 3 Automated Cell Counter was used to count cells for calculations of specific seeding densities. All counts were performed in duplicate. Countess 3 returns both total and live cell values, and the mean of the live values was used in calculations, while the percentage of live cells was used to assess cell health.

3.6.4. Freezing cells

ASC52telo cells purchased from ATCC were expanded and frozen at a low passage. To freeze cells, the cells were harvested, counted and resuspended in growth medium containing 10% (v/v) DMSO at a concentration of 500,000 cells/mL. Aliquots of 1.0 mL were transferred to 1.5 mL Cryotubes and placed in a CoolCell Container (Corning) at -80 ºC over night for gradual freezing, before being transferred to liquid nitrogen for long-term cryopreservation.

3.6.5. Thawing cells

Single aliquots of cryopreserved cells were transferred from liquid nitrogen when needed and quickly thawed at 37 ºC for 1-2 min, and immediately diluted 1:10 in preheated complete medium. The suspension was centrifuged at 300 rpm for 5 min, and the DMSO-containing SN was removed. The cell pellet was resuspended in 10 mL complete medium and transferred onto a 10 cm cell culture dish. Once the cells had settled, the culture dish was placed in the cell incubator.

3.7. Genotyping ASC52 cells

3.7.1. Genomic DNA isolation

The rs1733339 alleles of ASC52telo were genotyped by Chelex extraction of genomic DNA and subsequent sequencing of the locus. Chelex-100 resin binds divalent metal ions, thus chelating/sequestering them away from DNA-degrading nucleases. In addition Proteinase K and RNase A were added to digest cellular proteins and RNA, respectively (Walsh et al., 2013).

15 000 ASC52telo cells were grown in each of two wells of a 96-well plate for 24 h before genomic DNA was extracted; cells were washed once in 1X PBS, then 50 µL Chelex solution (0.5 % (w/v) Chelex-100 powder, 200 µg/mL Proteinase K and 50 µg/mL RNase A in ddH20) was added to each well. The mixture was pipetted up and down and scraped to aid cell lysis, then incubated 2 min at RT to allow for RNase action, followed by 4 h incubation at 56 ºC while shaking (210 rpm) for Proteinase K digestion. The plate was then spun down to collect evaporation, and complete lysis was ensured by gently pipetting up and down multiple times. The samples were transferred to 0.2 mL PCR tubes and centrifuged at 600 x g for 2 min to pellet the Chelex beads. SN (35 µL) was transferred to fresh tubes, and DNA concentration was quantified by QIAexpert. Cell-less, but equally treated Chelex solution used as blank.

Template for sequencing was then generated by PCR amplification of the region surrounding rs1799993 from the extracted genomic DNA. A PCR reaction with 10 ng gDNA, and primers 643F and 644R (Table 2.4.3) was run as described in Section 3.4.1, followed by agarose gel electrophoresis (1.5 %) of a small sample/aliquot to verify correct size of band lacking unspecific amplification products (Section 3.4.2). Enzymatic clean-up was run on the remaining PCR product sample to degrade DNA polymerase and unused dNTPs: 2 µL Illustra ExoStar was mixed with 5 µL PCR product and incubated for 15 min at 37 ºC, followed by deactivation at 80 ºC for 15 min. PCR amplicon concentrations were determined by NanoDrop. 100 ng of each sample was used as template for BigDye sequencing using the same primers as for template PCR (Section 3.4.4) and subsequent analysis via SnapGene.

3.8. Transfection optimization of MSCs for Prime Editing

The prime editing vectors were non-viral, hence a series of transfection tests was run to optimize the delivery of plasmid DNA (pDNA) to MSCs. A selection of transfection reagents were tested (Table 2.11), as well as electroporation and nucleofection methods. Most tests were conducted on the target cell line ASC52telo, but early in the project before these had arrived, the A41WAT preadipocytes were used as proxy, as they show similar sensitivity to transfection. Transfection in HEK293T cells was performed to control for plasmid function, as this cell line is easily transfected and has high capacity for transgene overexpression (Berk, 2005). Cells were plated 18-24 h prior to transfection at a density that produced 50-80% confluency at time of transfection. All transfection tests (unless otherwise specified) were performed in 96-well plates with

100 µL medium per well and cell seeding densities ranging from 3000-7000 cells/well. Nuclight RapidRed DNA dye was used in some cases as a live-cell label to quantify cell density. In these cases, Nuclight RapidRed reagent was added to the cell suspension directly prior to seeding at a 1:1000 (v/v) concentration, and cell suspension, as well as plated cells, were kept in the dark until imaging.

Transfection efficiencies were assessed by transfection of GFP-containing plasmids of various sizes (Table 2.7, materials), followed by monitoring of GFP expression using live cell fluorescence microscopy. Similar plasmids not containing GFP and/or non-transfected cells were used as negative controls. For most assays, the automated high-throughput Incucyte S3 system (Sartorius) was used to take pictures and quantify the number and intensity of fluorescence. For indicated assays, manual fluorescence microscopy was performed using the Nikon Ti2000 Eclipse microscope due to technical issues with the Incucyte.

3.8.1. Live-cell monitoring of transfection efficiencies

Most transfection experiments were monitored using the Sartorius IncuCyte S3 Live-Cell Analysis System (optical module S3/SX1 G/R; Table 2.3), where cells were imaged using the 10x objective with phase contrast, green fluorescence (300 ms capture time), and (when applicable) red fluorescence (400 ms capture time). For 96-well plates, 4 images were taken per well, thus capturing the majority of the surface area. For 6-well plates, 121 images were taken per well.

To quantify transfection efficiencies, analyses were carried out using the Incucyte software (Incucyte 2021C; Table 2.2). For each experiment, an individual analysis definition was built using the Basic Analyzer. To do so, parameters are adjusted to generate masks that covered the object in question (e.g. covering all cells in phase contrast images to quantify confluency, or covering fluorescent signals to quantify area and number of fluorescent objects). Analyses were trained on selected images, then applied to the entire image set for quantification. For confluency masks, adjustable parameters included segmentation adjustment (background sensitivity), hole fill (fill in holes under a certain size, if such holes were occurring on cells), pixel adjustment of filter size (adjust filter edge), eccentricity (specifically include or exclude objects that are exceptionally round or exceptionally elongated), and minimal or maximal requirements for an object to be included (e.g 2000 μm^2 minimum for ASC52-cells). For masks quantifying fluorescent objects (e.g. GFP+-expressing cells), in addition to parameters also used for confluency, adjustments included background fluorescence subtraction (Top-Hat or Surface-Fit segmentation algorithms), edge sensitivity (separate adjacent objects), and mean intensity levels (e.g. separates GFP-fluorescence from autofluorescence). Output was generated as averages per image for each well.

3.8.2. Chemical methods of transfection

Transfection reagents tested included two polymer-based reagents, Xtremegene 9 and TransIT-X2; two liposomal reagents, Lipofectamine 2000 and Lipofectamine Stem Reagent; and one non-liposomal lipid/protein/polyamine-based reagent, TransIT-LT1 (Table 2.11). All transfection reagents were assessed

with one or more of 4 different pDNA concentrations and two or more DNA (μg)-to-transfection reagent (μL) ratios ranging from 1:1 to 1:5 (Table 3.2). All reagents were used for forward transfection as described by manufacturer, with specifics as outlined in Table 3.2. In general, the pDNA-reagent complexes were prepared by mixing desired amounts of the two components with a specified amount of OptiMEM (at RT) and allowing for complexes to form for the amount of time specified by manufacturers. During complex formation, medium was changed on the cells. The DNA-reagent complexes were then added dropwise to the cells, and cells were returned to the incubator.

Lipofectamine Stem Reagent was tested as reverse transfection in addition to standard forward procedure. In reverse transfection, the pDNA-reagent complexes were added to the cells in suspension directly prior to seeding, rather than to attached cells one day after seeding. Cell seeding number was adjusted up to account for the lack of 24 h proliferation between seeding and transfection. Other than that the two protocols were identical.

Table 3.2 | Transfection optimization for Prime Editing of ASC52telo cells.

Cell line (cells/well)	**pDNA** (ng/well)	**Reagents**	**Ratios**	Cas9n-EGFP	eCas9-GFP	PE2-GFP	pmax GFP	pCMV5-empty	pU6-GG	Non-T control
A41WAT										
6000	40 80 120 160	Xtreme-Gene 9	1:3 1:5	x	-	-	-	-	x	x
6000	40 80 120 160	TransIT X2	1:3 1:5	x	-	-	-	-	x	x
3000 6000	120 160	TransIT X2	1:1 1:2 1:3	x	-	-	-	x	-	x
ASC52										
7000	120	TransIT X2	1:2 1:3 1:5	x	-	-	-	x	-	x
7000	120	TransIT LT1	1:2 1:3 1:5	x	-	-	-	x	-	x
7000	120	LF 2000	1:2 1:3 1:4 1:5	x	x	-	-	x	x	x
Fwd: 5000 Rev: 7000	60 120 160	LF Stem Reagent	1:2 1:3 1:4	-	x	x	x	-	-	x
HEK293T										
7000	120	TransIT X2	1:2 1:3 1:4	x	-	x	-	x	x	x
7000	120	TransIT LT1	1:2 1:3 1:4	x	-	x	-	x	x	x

Cas9n-EGFP, pDG461-Cas9n(D10A)-2A-EGFP; eCas9-GFP, pCAG-eCas9-GFP-U6-gRNA; PE2-GFP, pCMV-PE2-P2A-GFP; pU6-GG, pU6-pegRNA-GG-acceptor-empty-RFP; Non-T, non-transfected. LF, Lipofectamine.

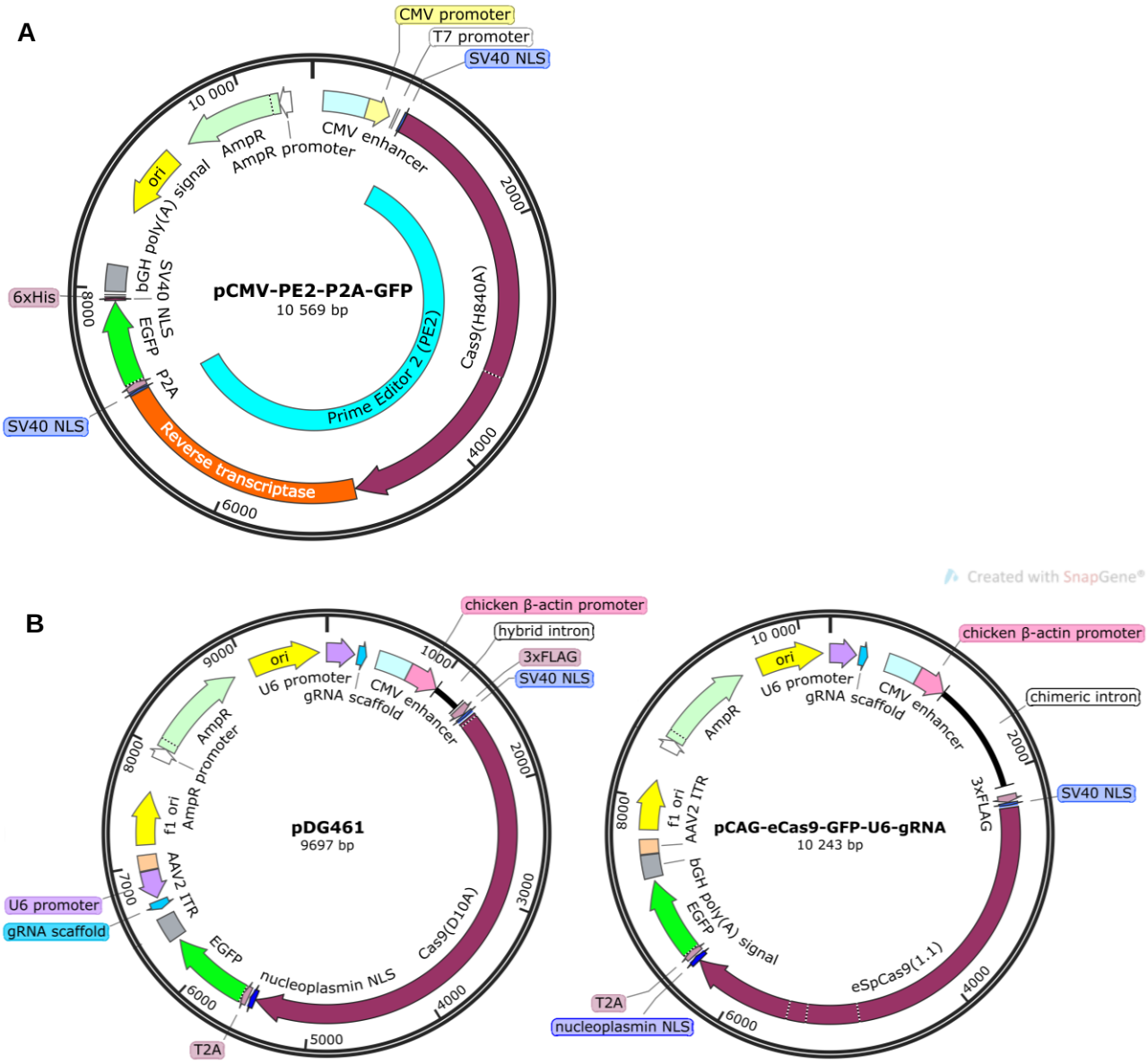


Figure 3.5 | Vector maps of plasmids used in transfection optimization of ASC52telo. A) The target plasmid, pCMV-PE2-P2A-GFP, and **B)** two surrogate-plasmids, pDG461-Cas9n(D10A)-2A-EGFP (left) and pCAG-eCas9-GFP-U6-gRNA (right).

3.8.3. Electroporation

Electroporation is a physical transfection method where an electrical current is generated to open pores in the cell membrane and transport the DNA molecules into the cell cytoplasm. This method was performed on A41WAT cells using the BioRadGene Pulser Xcell Electroporation system and accompanying cuvettes (0.4 cm gap), and the plasmid pDG461-nCas9-EGFP (Table 2.7).

Cells were harvested at 80 % confluency, counted, and adjusted to 1.5 million cells/ml in OptiMEM. Cuvettes and cell suspension were kept on ice prior to electroporation. For each reaction, 100 µL cell

suspension was mixed with variable amounts of plasmid DNA (2.5 μg, 5.0 μg, 10.0 μg, or 25.0 μg) in the cuvette and pulsed at one of four different voltages each (160 V, 190 V, 220 V, or 250 V), such that a total of 16 different electroporation conditions were tested. Following electroporation, 900 μL pre-heated (37ºC) complete proliferation medium was added gently to the cuvette (not mixed), and cells were incubated at RT for 5 min. 100 μL reaction mixture was then aliquoted to each of 6 wells in a 96 well plate and incubated for another 20 min at RT before being placed in the IncuCyte cell incubator for monitoring of cell growth, viability and GFP expression for 72 h.

3.8.4. Nucleofection

Nucleofection is an advanced method of electroporation that uses a specialized Nucleofection device for delivery of cell type-specific electrical pulses to cells incubated in an optimized nucleofection solution with the plasmid DNA. The result is that pDNA molecules are transported directly into the cells nucleus in stead of, or as well, as to the cytoplasm (Nakashima et al., 2005). This method was applied to the ASC52telo cells and was run on the 4D-Nucleofector System using the P1 Primary Cell 4D-Nuclefector X Kit. The kit uses a 16 well-Nucleocuvette Strip with 20 μL reaction mixture per well, 50k cells per reaction. For mesenchymal stem cells, the P1 program is recommended with one of two different pulse conditions; EW-104 for high viability, and FF-104 for high transfection efficiency. To assess which of the two pulse conditions resulted in the best results for ASC52telo, a preliminary test was run as suggested with the accompanying pmaxGFP plasmid at 4 different concentrations on both pulse conditions in parallel.

The cells were harvested directly prior to nucleofection and counted. An aliquoted number of cells was centrifuged again at 300 x g, 5 min, and SN removed. Cells were resuspended in Nucleofector Solution with Supplement (mixed in a ratio of 1:4.5), to yield a concentration of 50 000 cells/20 μL. 20 μL aliquots were mixed with variable amounts of pmaxGFP plasmid (0 ng, 200 ng, 400 ng, or 800 ng in 1.6 μL TE-buffer) and transferred to the Nucleocuvette strip and immediately nucleoporated. Following pulse completion, the cells were rested for 10 min at RT before the addition of 90 μL pre-warmed medium. From each diluted suspension of nucleoporated cells, 25 μL (11.4k cells) was aliquoted to each of 4 replicate wells of a 96-well plate, prefilled with 75 uL medium. The same number of non-nucleoporated cells were seeded as controls. The plate was incubated overnight in the cell incubator, then imaged with Nikon Ti2000 Eclipse fluorescence microscope.

The preliminary test showed that the program EW-104 had the highest viability and transfection efficiency for ASC52telo cells. Thus, a second test was run with complete Prime Editing set-up in 6-well plate format, with the target plasmid pCMV-PE2-P2A-GFP co-transfected with the various pegRNA-encoding plasmids (with inserts, Table??) as well as the ngRNA. The procedure was the same as above, using the highest total amount of plasmid with the following ratios; for each reaction of 50 000 cells, 600 ng PE2-GFP plasmid was mixed with 150 ng pU6-pegRNA-GG, and 62 ng ngRNA (A) plasmid (ratios based on (Anzalone et al.,

2019)), for a total of 6 plasmid combinations, as well as one with only 600 ng pCMV-PE2-GFP plasmid. Two replicates were prepared for each plasmid combination, and both were run with the pulse program EW-104. The two replicates were combined in a single well of a 6-well plate, for a total 100 000 cells/well. Controls included plasmid-free nucleoporation, as well as two non-nucleoporated samples where one was incubated in nucleofection solution for 15 min prior to seeding, and the other not. The cells were monitored for transfection efficiency and viability with the IncuCyte Live-Cell Monitoring System.

3.9. Enhancer CRISPR-mediated activation/inhibition

3.9.1. Overview of the CRISPR activation/inhibition workflow

In enhancer CA/I, the DNA of the target genomic locus is not edited. Instead, a catalytically dead Cas9 (dCas9) is exploited to bring epigenetic remodelers and transcriptional coregulators to the locus to either activate or repress the enhancer and the downstream transcription of the target gene(s) (Figure 1.6, introduction) (Li et al., 2020). To achieve stable expression of dCas9 and the epigenetic remodelers, lentiviral transduction was performed, which is a method of gene transfer where the gene of interest is stably integrated into the host genome at a non-specific location via reverse transcription to generate a stable cell line (Durand & Cimarelli, 2011). Moreover, to avoid bias on cell survival or function during establishment of the stable cell lines, and, more importantly, to be able to manipulate the enhancer in various stages of mesenchymal stem cell differentiation, the system was made inducible by including a Tet-on-3G/doxycycline control system (Figure 3.6, left insert). Briefly, the (*p300- or KRAB-*)dCas9 fusion constructs are under control of the TRE3G promoter, which can only be activated by the transcriptional activator Tet-on-3G in the presence of doxycycline. The Tet-on-3G sequence is thus separately integrated into the genome under a constitutive promoter (pSFFV), and doxycycline is added to the cell medium at the desired time point to induce expression of the CA/I system (Figure 3.6, Steps 1-2). By tagging the expression cassettes with different fluorescent markers, cells containing both constructs can be identified and sorted by flow-assisted cell sorting (FACS) (Figure 3.6, Step 3). Following sorting, doxycycline is removed form the medium, resulting in cells poised to express the dCas9 fusion proteins. Finally, the sgRNAs, as well as the transcriptional coregulator (VP64 activator or LSD1 repressor), encoded on a third lentiviral vector with a third fluorescent marker, can be transduced to complete the system (Figure 3.6, Steps 4-7). While this expression cassette is stably expressed in the target cells, no manipulation of the enhancer occurs until dCas9-p300/KRAB is induced by doxycycline (Figure 3.6, Step 8) (Li et al., 2020).

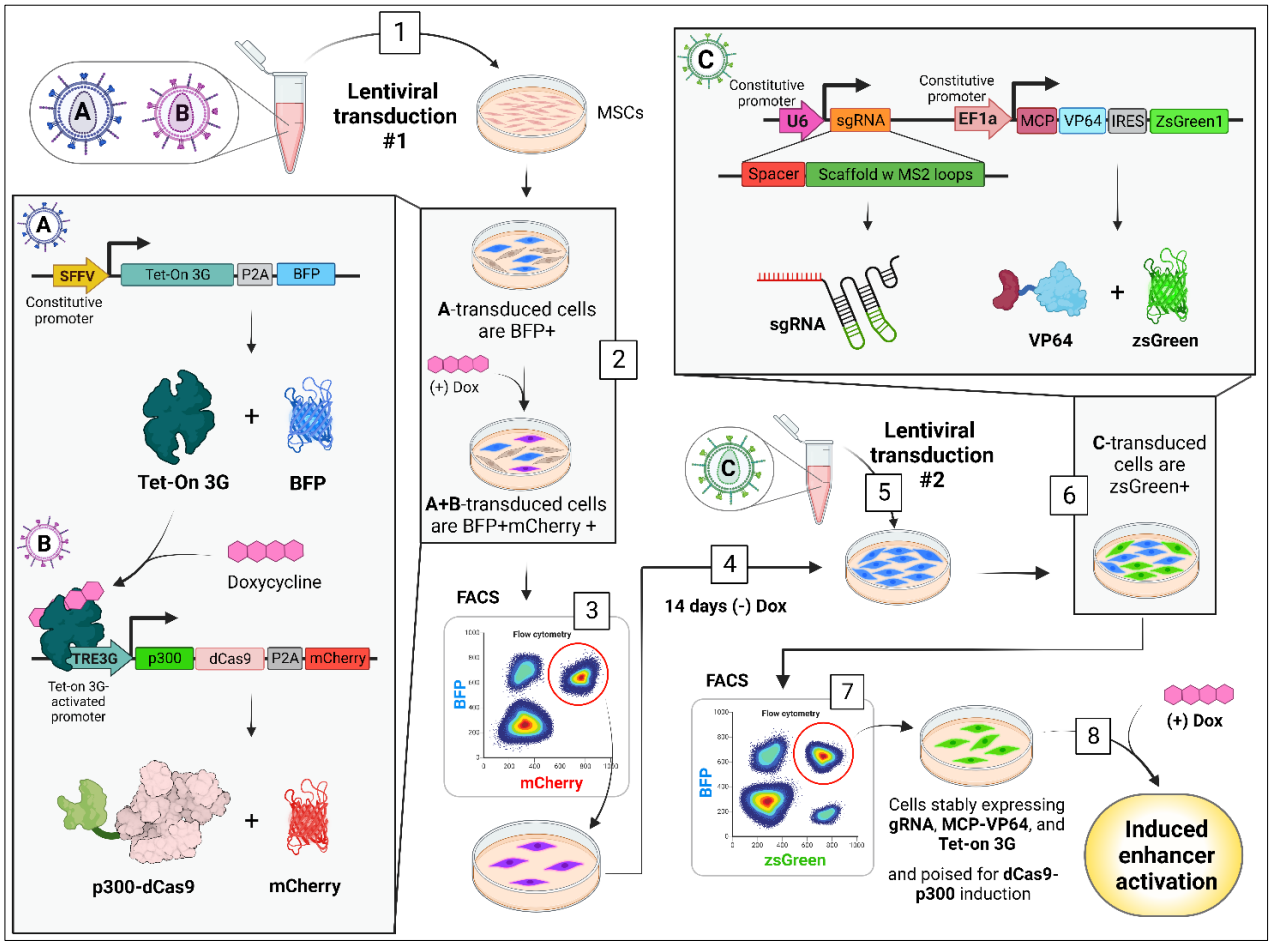


Figure 3.6 | Workflow for establishing an inducible CA/I MSC cell line. (Details are shown for enhancer *activation*, but the principles are identical for enhancer *inactivation* except for the specific effectors involved [p300 and VP64 vs LSD1 and KRAB, respectively]). **Step 1:** Dual lentiviral transduction of MSC's using viral particles with genomes A and B. **Step 2 and left insert:** Cells with integrated transfer sequence A express the bicistronic Tet-on-3G-BFP sequence under the control of constitutive promoter SFFV, while cells transduced with transfer sequence B receive the bicistronic effector(p300)-dCas9-mCherry sequence. However, since the latter sequence is under the control of the Tet-on-3G-dependent promoter TRE3G, B-transduced cells will only express mCherry if also transduced with A and cultured in the presence of doxycycline. **Step 3:** Thus, A+B-transduced cells can be selected for via FACS for mCherry alone. **Step 4:** After expanding for 14 days with no doxycycline, the doubly transduced cells will no longer be expressing mCherry, and **(Steps 5 and 6:)** are ready for second round of transduction with viral transfer-sequence C containing the sgRNA(MS2) sequence as well as the MCP-TF(VP64)-zsGreen1 sequence **(insert top-right)**, both under control of constitutive promoters. **Step 7:** Again, cells are sorted for successful transduction via FACS, and expanded. At this point cells are stably integrated with all three transfer-sequences, and are constitutively expressing sequences A and C, but not sequence B due to no doxycycline present in medium. (Presence of MCP-TF has no effect on transcription when not tethered to DNA). **Step 8:** Doxycycline can be added to medium at time of interest to induce enhancer modulation (e.g. at any stage of differentiation) for assessment of changes in downstream gene expression. *Method adapted from (Li et al., 2020), figure made in Biorender.*

3.9.2. Lentivirus particle production in HEK293T cells

Before performing the workflow outlined above, the lentiviral vectors had to be prepared. First, the sgRNAs were cloned into vector (C) and all vectors were amplified and purified as described in Section 3.5. Next, to produce viral particles, the envelope and packaging plasmids were transfected into a packaging cell line, HEK293T, together with the transfer plasmid (vector A, B or C, Figure 3.7). Each transfer plasmid was packaged in a separate reaction. HEK293T cells were seeded 18-24 h prior to transfection in 10 cm dishes at a density that produced 50-80% confluency at time of transfection, in antibiotic-free culture medium. Cells were transfected with Lipofectamine 2000 according to manufacturer's protocol, with 6 µg VSV-G plasmid, 12 µg Gag-Pol plasmid, and 12 µg transfer plasmid, for a total of 30 µg plasmid per transfection, using 1:2 DNA-to-reagent ratio. Cell medium was changed directly before transfection (antibiotic-free), and again 7-18 h after (with antibiotics). Virus was harvested at 48 h (first harvest) and 72 h (second harvest) after transfection: Cell medium containing virus was collected and either frozen directly at -80 ºC (for virus containing transfer sequences A and B; Figure 3.7), or flash frozen in liquid nitrogen before long-term storage at -80 ºC (transfer sequences C; Figure 3.7). The two harvests were not combined. Following the first harvest, the collected media was replaced with fresh (also with antibiotics).

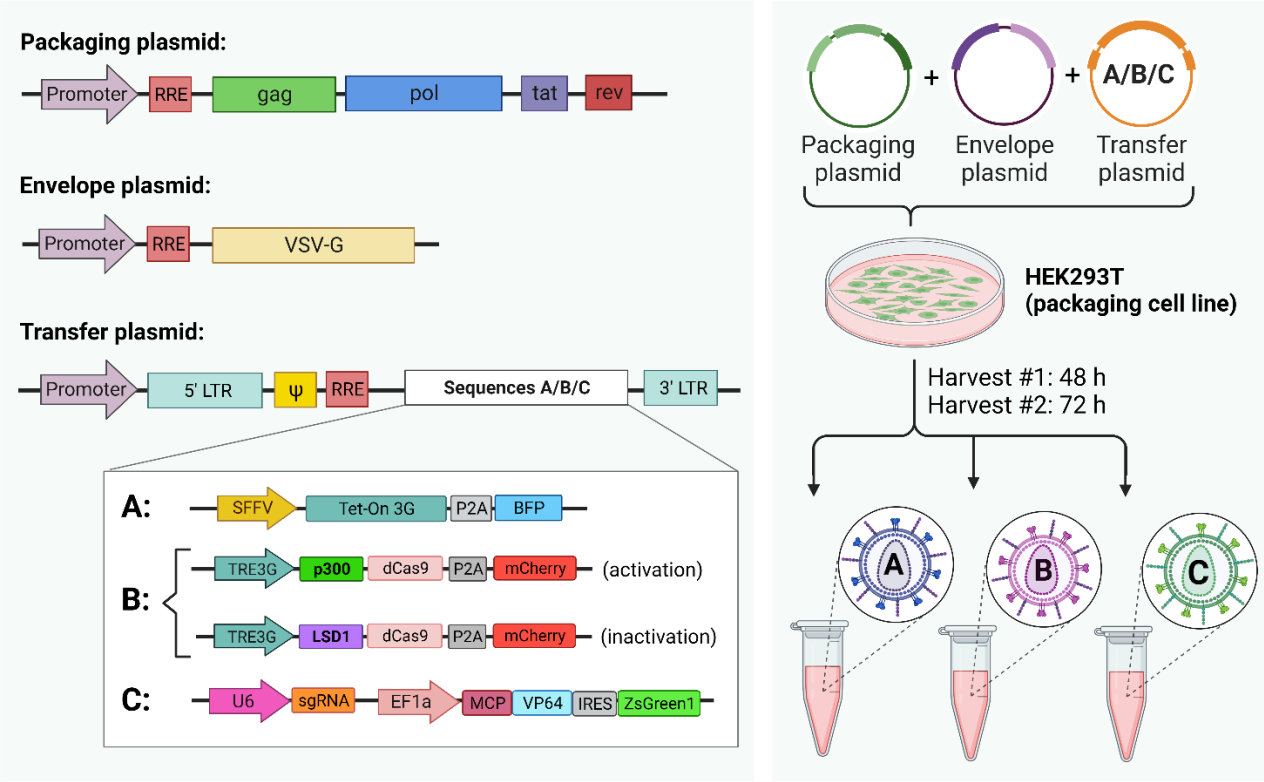


Figure 3.7 | Lentiviral packaging in HEK293T cells. The lentiviral particles were produced in the packaging cell line HEK293T, transfected with 2nd generation lentivirus packaging plasmids. For the 2nd generation system, the lentiviral HIV-1 genes have been divided between two plasmids; the envelope plasmid and the packaging plasmid, thus reducing the risk of producing replication-competent viral particles. The packaging plasmid contains the *gag-pol*

genes, as well as the two regulatory genes *tat* and *rev*. *gag* encodes the capsid, nucleocapsid, and matrix proteins which form the virion capsule, while *pol* encodes the proteins responsible for integrating the transfer gene into the host genome. The envelope plasmid encodes the proteins for surface proteins that determine the cell tropism (target-cell specificity). Most lentiviral systems (including that used in this thesis) have been pseudotyped with the vesicular stomatitis virus *VSV-G* envelope gene, to confer broader tropism. The transfer plasmid contains the transfer sequences to be integrated into the host genome, as well as several elements necessary for incorporation into virions (e.g. the psi packaging signal and RRE). The transfer sequence is flanked by the 5' and 3' LTRs (long terminal repeats), which mediate integration into host genome by integrase. **Left panel** shows the essential elements of the three plasmids required for each packaging reaction, while **right panel** depicts the general packaging reaction in HEK293T. The lentiviral particles were harvested at 48 h and 72 h post-transfection. Each transfer sequence was packaged in a separate reaction. (For maps of plasmids harboring sequence A and B, see Figure 3.8, and Figure 3.4 C for sequence C. Plasmid maps for packaging and envelope plasmids not available.) *Figure prepared in Biorender.*

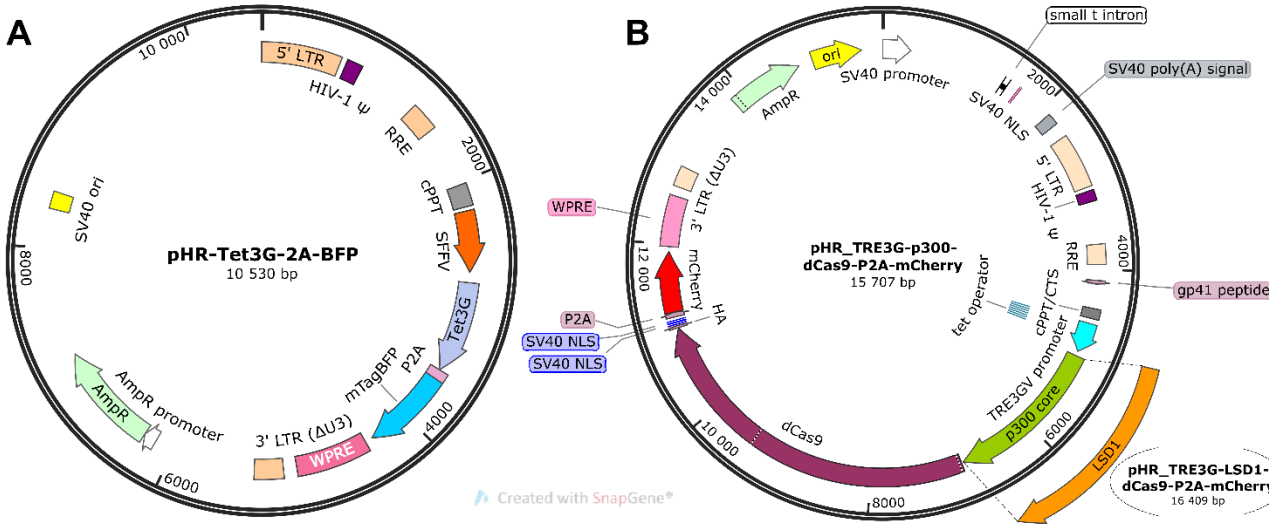


Figure 3.8 | Plasmids used with lentiviral transduction. A) pHR-Tet-on-3g-BFP. B) pHR-TRE3G-p300-dCas9-P2A-mChery, with pHR-TRE3G-LSD1-dCas9-P2A-mChery insert. Plasmids sourced from (Li et al., 2020).

3.9.3. Up-concentrating viral stocks

Virus-containing medium was thawed on ice after 1 month of storage at -80 ºC, then centrifuged at 300 x g, 5 min, to remove cells and debris. SN was removed and filtered through a 0.45 µm polyethersulfone syringe-filter, then mixed with 1 volume Lenti-X Concentrator per 3 volumes virus-containing SN and incubated at 4 ºC for 45 min. Following incubation, the mixture was centrifuged at 1500 x g, 45 min, 4 ºC, to pellet the virus. SN was carefully removed and discarded without disturbing the pellet. Pellet was resuspended in 1/20th original volume of target cell (ASC52telo) medium (4 ºC), such that the virus had been up-concentrated 20X. The virus was further quantified and used directly.

3.9.4. Quantification of virus particles

Concentrated virus stocks were quantified for viral particle count by the use of Lenti-X GoStix Plus. Lenti-X GoStix are single-use antibody tests that determine the concentration of viral capsid protein p24, which can then be converted to a virion particle number to determine titer. The tests had a cassette where the sample is applied, and a test-window where a control band (C) and a test band (T) appear. 20 µL concentrated virus stock was added to the sample well (S) of the GoStix cassette, followed by 80 µL Chase Buffer. After 10 min incubation, the result was analyzed by the use of an accompanying app downloaded to a mobile device: The camera was used to scan the test result, and the app converted the intensity of the band to "Go value (GV)", which is equal to ng/mL p24, by way of a lot number-specific standard curve premade by the manufacturer. If the reading exceeded the standard curve and an "Off scale" value was returned by the app, an aliquot of the stock was diluted 1:10 and retested.

Given the molecular weight of p24 and the knowledge that there are about 2000 p24 molecules per physical viral particle (PP) (German Advisory Committee Blood, 2016), 1 ng p24 contains about 10 million PP. The number of PP capable of infection (infectious units (IFU)) varies depending on multiple factors, but typical values range from 1/1 to 1/200 (Sigma Aldrich, n.d), and is usually estimated from a titer assay before transduction. Due to time constrains, we estimated an intermediate packaging efficiency for the plasmids and cell line involved, of 1/80 IFU/PP. Thus, to calculate viral titer, the p24 concentration was converted to theoretical IFU concentration by the following equation: x ng/mL p24 * 10^7 PP/ng * 1 IFU/80 PP = x IFU/mL, (PP=physical particles, IFU = infectious units). For transduction, different amounts of IFU-to-target cell ratios (commonly known as Multiplicity of infection, MOI)), ranging from 1 to 30, were tested.

3.9.5. Lentivirus transduction of ASC52 cells

Lentivirus transduction was conducted for both enhancer CRISPR *activation* (CA) and *induction* (CI) in parallel, such that cells received virus containing either activating or inactivating CA/I sequences; p300-dCas9-mCherry for enhancer activation, and LSD1-dCas9-mCherry for inactivation. Both series received the Tet-on-3G-BFP-containing virus for doxycycline-induced expression of the dCas9-sequence. ASC52telo cells were seeded in 6-well plates 24 h prior to induction, at 80 000 cells/well (= 8 300 cells/cm^2), and estimated to be about 100 000 cells at the time of transduction (for MOI calculations). 4 MOIs were chosen for the initial transduction; 1, 5, 10 and 30. To activate integrated transgene expression, the cells were incubated in medium containing doxycycline (dox, 1 µg/mL) during and following induction. Control cells were either not transduced (no virus), or transduced with the highest MOI (=30) but did not have transgene expression induced (dox-free). Cells were transduced in complete medium containing 10 µg/mL Polybrene, 1 µg/mL doxycycline (except no-dox controls), and the required volume of the respective virus preparations of Tet-on-3G and (p300-dCas9 or LSD1-dCas9) for the desired MOIs. To enhance virus-target cell contact, the volume of medium was

reduced to half of normal during time of transduction. Virus-containing medium was removed and replaced with fresh medium (with or without doxycycline) 16 h after transduction.

3.9.6. Flow cytometry of transduced ASC52telo cells

The transduced cells were assessed by flow cytometry using the Sony SH800S Cell Sorter. Because this flow cytometer does not have the ability to detect the BFP marker on the Tet-on-3G construct, the downstream mCherry marker on the dCas9-p300/KRAB, which should only be expressed in the presence of Tet-on-3G, was used to assess successful doubly-transduced cells.

CA-transduced cells: cells were harvested at 72 h post-transduction, and following centrifugation, each pellet was resuspended in 1X PBS and centrifuged again, 300 x g, 5 min. Cell pellet was resuspended in 500 µL PBS containing 1:1000 LIVE/DEAD Fixable Green Dye (= 1/250 000 AU final concentration) and incubated for 30 min in the dark. Cells were then pelleted (300 x g, 5 min), then washed twice by resuspending in FACS buffer and pelleting between. **CI-transduced cells:** cells were harvested and washed twice in FACS buffer. No LIVE/DEAD marker was used. Directly prior to sorting, all cells were vortexed and filtered through a 100 µm cell strainer to ensure a single-cell suspension.

For the Fixable green dye, the 488 nm laser was used for excitation, with the FL1 (525/50) filter to detect emission, while mCherry was excited with the 561 nm laser and detected through the FL3 (617/30) filter. Prior to sorting, transduced vs control cells were analyzed to set gating strategies: A series of four gates was set up; 1) forward scatter vs side scatter (FSC-A vs SSC-A) to remove debris, 2) a histogram of FL1-A filter to eliminate dead cells (for samples without live/dead marker, this gate was left open), 3) pulse geometry gating (area vs height; FSC-A vs FSC-H) to eliminate doublets, and finally 4) side scatter vs FL3-A filter to detect mCherry+ cells. Since no mCherry+-cells could be detected, sorting was aborted. Cells were instead checked for BFP on the BD FACSymphony S6 Cell Sorter by qualified personnel (using the 405 laser for excitation and 445 nm filter for detection). All cells were then replated in complete medium and expanded for two weeks, then frozen for future analysis. Flow cytometry data was analyzed with FlowJo v10.9.

4. Results

4.1. sgRNA design for Prime Editing and enhancer activation/repression

Since both the major methods used in this thesis are based on the CRISPR/Cas9 system (PE and CA/I), the first step was to design spacer sequences that would successfully direct the Cas9-fusion proteins to the target locations. For Prime Editing (PE), the Cas9-based Prime Editor needed to locate in close proximity to the SNP to be edited (rs1799993), while for CA/I we wished to select a variety of spacers that would cover the proposed enhancer region, termed Tile 8a (Figure 4.1 (top panel)). For PE of rs1799993, the PrimeDesign software suggested the same spacer that had previously been manually designed as the top choice due to the

ideal position relative to the SNP to be edited (Spacer-01, Figure 4.1 (bottom panel), and Table 4.1). However, a second spacer (Spacer-02) was also identified, and although situated further from the SNP, it was still within the 35 bp editing window, and demonstrated consistently superior predicted efficiency scores across several methods of assessment (Figure 4.1 (bottom panel), and Table 4.1). PrimeDesign also suggested a spacer for use with the PE3b system that overlapped with rs1799993, and thus was expected to be selective for the one genotype over the other (Spacer-ng, Table 4.1)

For selection of CA/I, the guides were not restricted to a single SNP, but rather to the peak of the enhancer area in Tile 8a, covering a 700 bp window (Figure 4.1, top panel). The UCSC genome browser track "CRISPR-targets" revealed a large number of high-scoring spacers (green color) in this window (Figure 4.1, bottom panel). Among these, we identified 3 spacers with consistently high predicted efficiency scores across several assessment methods (Figure 4.1, bottom panel and Table 4.1) in addition to the 2 spacers identified for PE, which could also be used for CA/I.

Considering that sgRNAs are predicted to be more efficiently transcribed from both the U6 and T7 promoters when the first nucleotide of the spacer sequence is a guanosine, for all sequences that did not naturally start with a G, a B-version was also designed for *in vitro* evaluation, in which the first nucleotide was changed to a G. Overall, the locus was found to be highly suitable for CRISPR activation/inhibition, as well as compatible with prime editing of rs1799993.

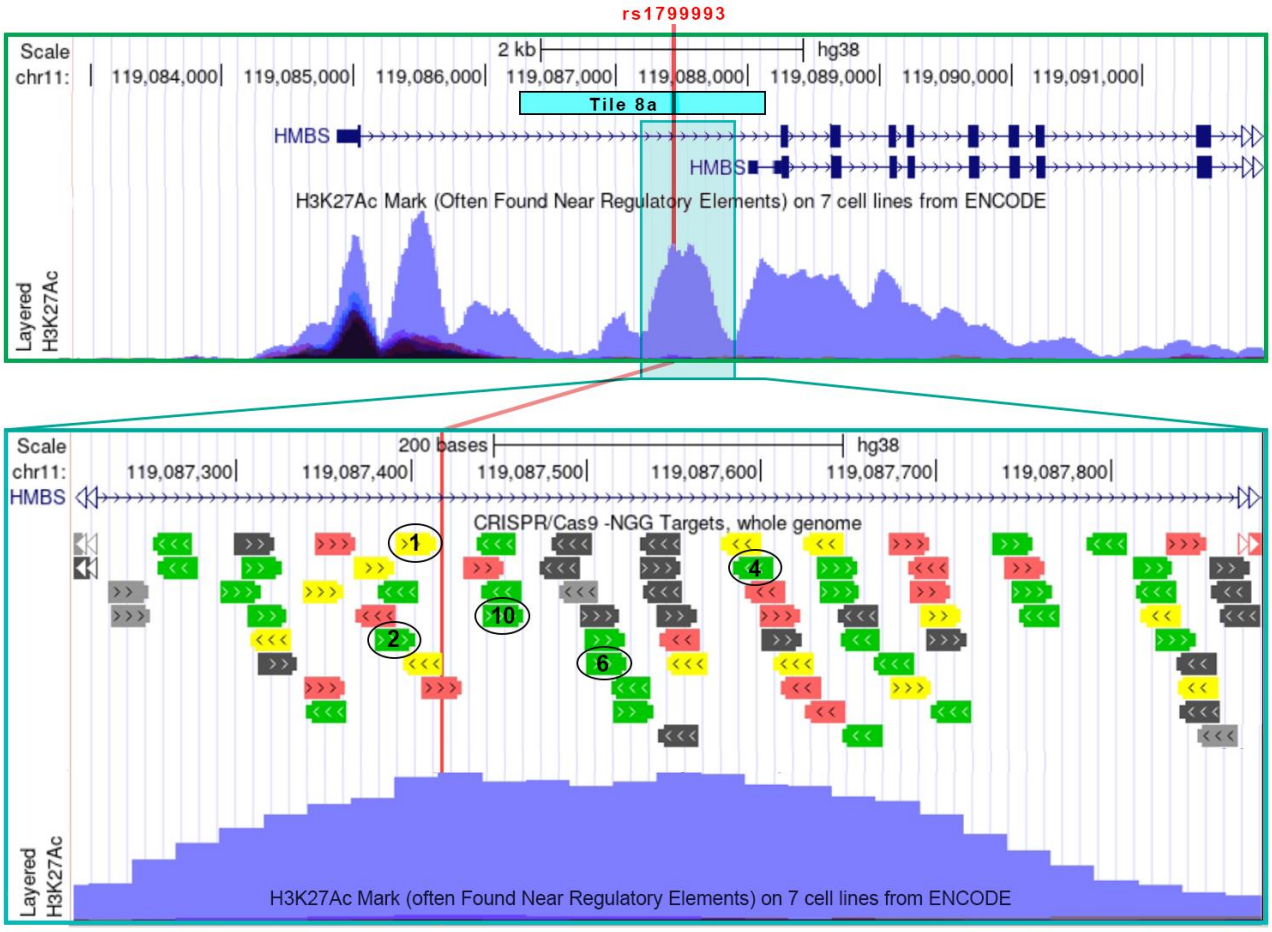


Figure 4.1 | Overview of best scoring guides for PE and CA/I. (Top panel) UCSC browser view of parts of the visceral obesity-associated 11q23.3 locus. Layered H3K27ac tracks show enhancer activity across multiple cell types. Tile 8a refers to a region previously experimentally shown to have genotype-dependent *in vitro* enhancer activity. (Bottom panel) Zoomed in view, with the "CRISPR targets" tracks shown. The position and predicted efficiency and specificity of possible spacers is shown. Green spacers, high specificity and efficiency; yellow spacers, high specificity but medium efficiency; red spacers: high specificity but poor efficiency: Black and grey spacers, not recommended due to low specificity. The best-scoring spacers are numbered and marked with black circles.

Table 4.1 | Spacer sequences for use with PE and CA/I targeting rs1799993

Spacer	Target system	MIT specificity #	Doench †	Moreno-Mateos ††	DeepHF †††	Distance* to rs1799993 (nt)	First nucleotide
Spacer-01	PE and CA/I	66	46 %	25 %	0.54	9	G
Spacer-02	PE and CA/I	68	92 %	96 %	0.67	21	T **
Spacer-04	CA/I	79	99 %	71 %	0.65	173	A **
Spacer-06	CA/I	84	86 %	88 %	0.67	115	C **
Spacer-10	CA/I	83	92 %	41 %	0.74	42	C **
Spacer-ng (C)	PE (PE3b)	76	46 %	52 %	0.53	15	G
Spacer-ng (A)	PE (PE3b)	76	46 %	52 %	0.53	15	G

The MIT specificity score ranges from 0-100. A score above 50 is recommended.

† The Doench *in cellulo* efficiency score is given as percentiles. Green tiles have a score above 55%, meaning that 55% of mammalian guides score equally well or lower than these guides.

†† The Moreno-Mateos *in vitro* efficiency score is given as percentiles.

††† The DeepHF efficiency score ranges from 0-1, where 1 is the best.

*Distance from cleavage site

** B-versions were prepared for spacer sequences not naturally starting with a G.

4.2. Screening of sgRNA cleavage efficiency on target DNA *in vitro*

The on-target cleavage efficiencies of the gRNAs containing the designed spacers were evaluated *in vitro* using the Guide-it *in vitro* Cas9 Cleavage Assay Kit from Takara Bio. For this assay, a target cleavage template that encompassed all the sgRNA target sites was generated by PCR from human genomic DNA (Figure 4.2). The assay was performed in duplicate on the same template region but generated from two different human samples; one homozygous for the risk allele of rs1799993, and the other for the protective allele. This was specifically to assess whether the genotype-specific (A) and (C)-variants of the ngRNA spacer would selectively cleave one template more efficiently than the other.

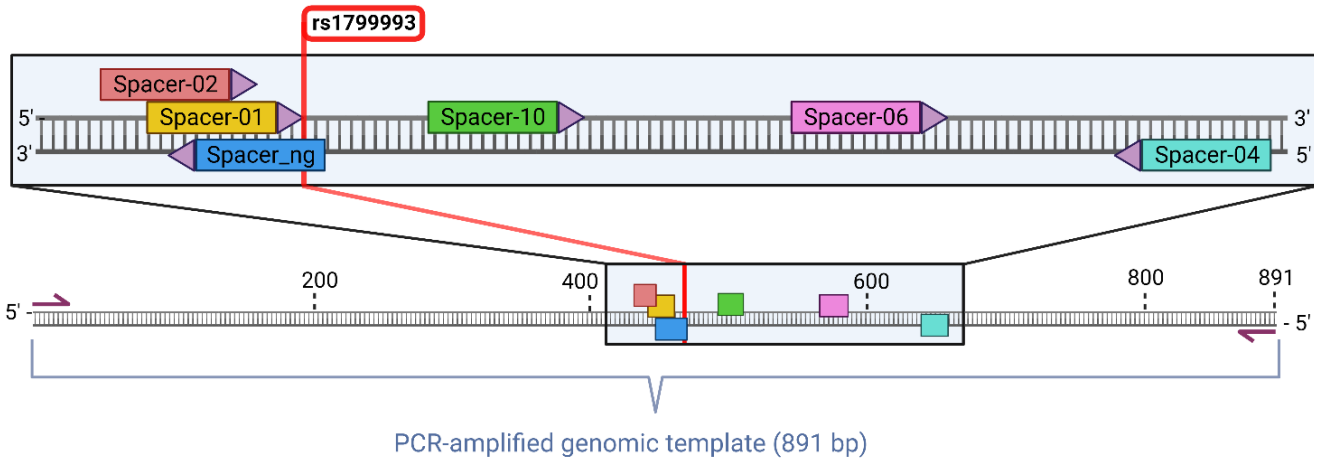


Figure 4.2. | Schematic diagram showing the relative locations of the chosen spacer sequences along the PCR-amplified template for in vitro cleavage assay. Adjacent PAM sequences are indicated with purple triangles. All spacers except Spacer-ng for use with CA/I, while spacers 01 and 02 were also suitable for use with PE targeting rs1799993 (red line). Two versions of Spacer-ng were prepared, specific for either C or A template, respectively, to be used with the PE3b system. *Figure made in Biorender*.

The spacer sequences were cloned into sgRNA scaffolds and transcribed *in vitro* by T7 polymerase. The purified sgRNAs were then assembled with recombinant Cas9, before the sgRNA-Cas9 complexes were added to the template. Following the cleavage reaction, the templates were analyzed on an agarose gel to assess how efficiently they had been cleaved (Figure 4.3. A). Successful cleavage would result in the disappearance of the full-length band, and the appearance of the cleavage products. Depending on the spacer target site, the cleaved template products were expected to be either of equal length (generating a single band), or of different lengths (generating two distinct bands).

All the tested sgRNAs resulted in Cas9 cleavage of the template and the generation of fragments of expected sizes (Figure 4.3 A). However, there were marked differences in the cleavage efficiencies of the various sgRNAs, and percentage of template cleaved ranged from 30-97 % (Figure 4.3 B). Overall, sgRNAs with spacer number 01, 02 and 10 performed best, with > 90% cleavage efficiency regardless of the template's genotype or the spacer's starting nucleotide. Interestingly, spacer-06 worked clearly best with the original configuration (A), while the (B) version, where the first nucleotide was replaced with a G, strongly reduced the efficiency. In contrast, spacer-04 consistently performed worse across all combinations, with mostly 50% efficiencies. Interestingly, there was no genotype-dependent difference in cleavage efficiency for the ngRNA spacer variants, despite the fact that they were specifically designed to be genotype-sensitive.

Taken together, both PE spacers worked excellently *in vitro* and were therefore both selected for subsequent testing *in cellulo*. For CA/I, the same spacers (01 and 02) as well as number 10 performed best and were chosen for subsequent testing *in cellulo.* However, because these 3 spacers are clustered closely together (Figure 4.2), the suboptimal spacers number 4 and 6 were also included as backups since they are situated further apart and thus may have a different effect on the enhancer activation.

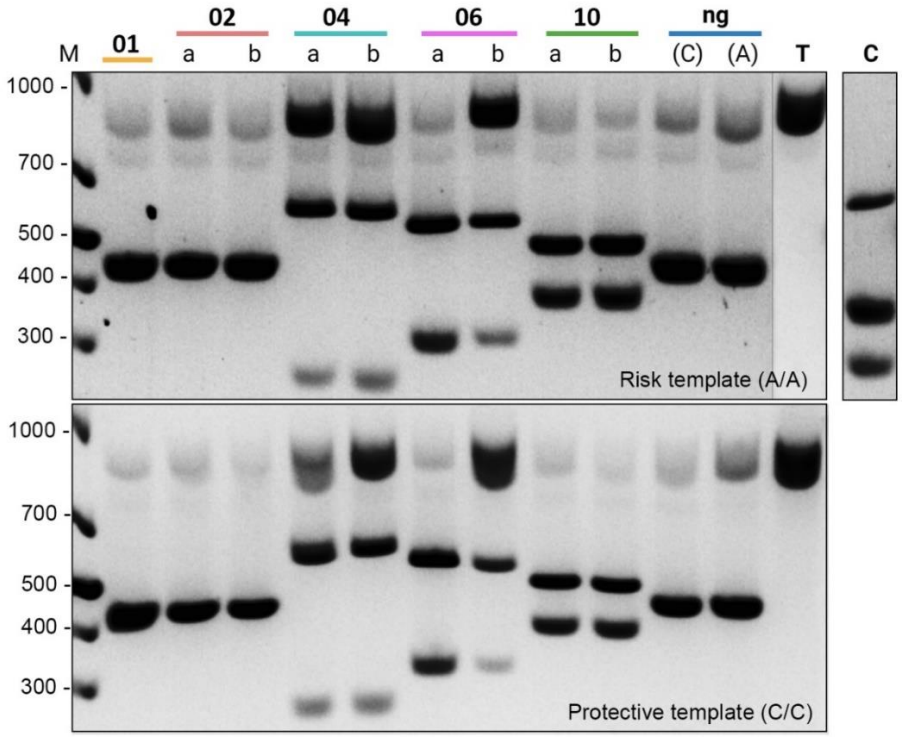


Figure 4.3 - part A | Guide-it *in vitro* sgRNA cleavage assay. sgRNA spacer sequences for targeting the 11q23.3 locus were evaluated *in vitro* for on-target cleavage efficiency. Purified sgRNAs were loaded onto recombinant SpCas9 and applied to the genomic cleavage template (891 bp), which was amplified by PCR from two human genome samples, one homozygous for the risk allele of rs1799993 (A/A), and the other for the protective-allele (C/C). **A)** Agarose gel verification of cleaved products. For each lane, band of ~900 bp indicates remaining uncleaved template, while shorter bands (variable lengths) indicate cleaved fragments. Lanes: M – marker. Numbers 01-10, and ng correspond to spacer sequence of sgRNA. T – undigested template control. C - control provided by kit. (Part B below)

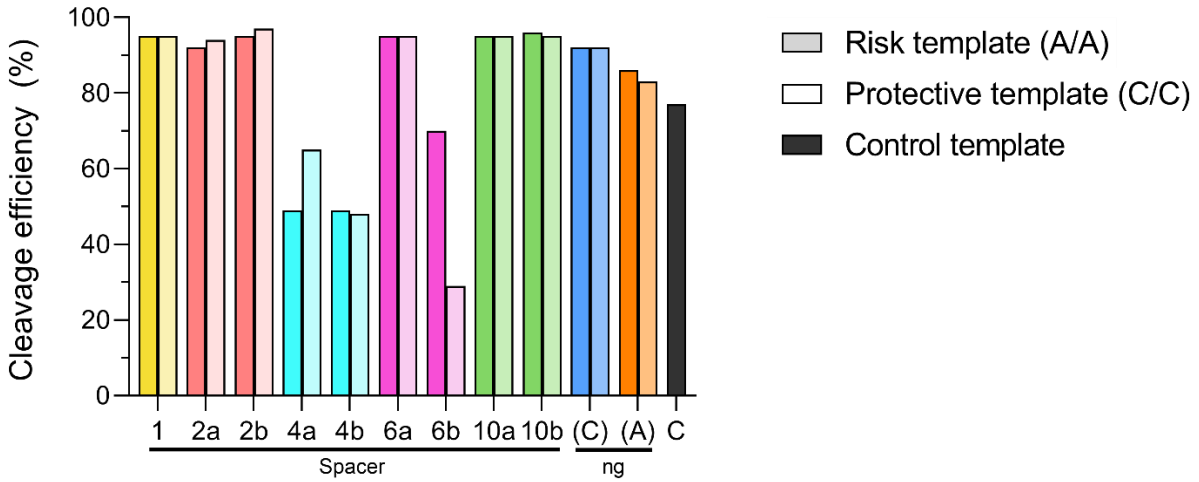


Figure 4.3 – part B | (see Figure legend part A). Cleavage efficiencies of the various sgRNAs (number 1 – 10, as well as the two ng-variants) were quantified based on the agarose gel of the results; densities of all bands per lane were measured and, combined densities of the cleaved products were calculated relative to total density of all bands. For each spacer sequence, the results for the two templates are displayed side by side (Risk template in darker color, and protective template in lighter).

4.3. Genotyping rs1799993 in ASC52telo

Before enhancer modulation could be performed in the target cell line ASC52telo, the genotype of the SNP rs1799993 had to be determined. The region harboring the SNP was therefore PCR amplified and sequenced with BigDye, which revealed that the ASC52telo cells were heterozygous at this locus (Figure 4.4).

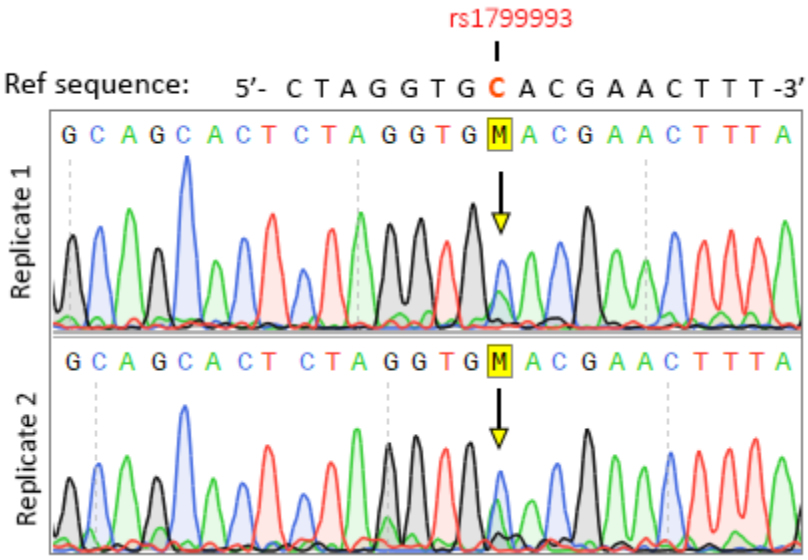


Figure 4.4. | Sequencing results for the genotyping of rs1799993 in ASC52telo cells. Genomic DNA was isolated from ASC52telo cells and sequences with BigDye. Both replicates showed two overlapping peaks for rs1799993 (yellow arrows); M = C (blue peak; protective allele) or A (green peak; risk allele).

4.4. Prime Editing of rs1799993 in ASC52telo

PE of rs1799993 was to be performed in the adipose-derived mesenchymal stem cell line ASC52telo to produce two cell lines that were genetically identical except for the genotype of the SNP rs179993, where one would be homozygous for the risk (A) allele, and the other for the protective (C) allele. Considering that the cells were discovered to be heterozygous for the SNP, PE would need to be performed in both directions simultaneously, to generate two homozygous ASC52telo cell lines; one each for the risk and protective alleles (Figure 4.5).

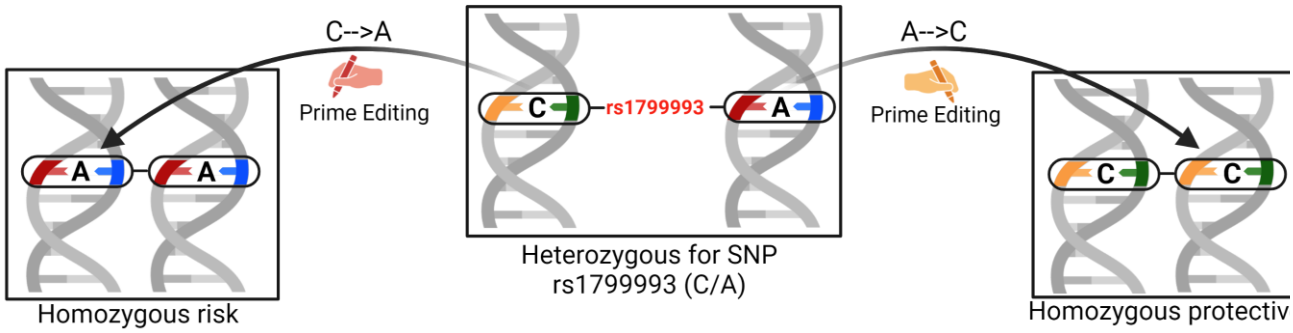


Figure 4.5. | PE strategy of the heterozygous ASC52telo cells to generate two homozygous cell lines. Genotyping of rs1799993 in ASC52telo revealed the WT cell line to be heterozygous at the locus, with the protective C-allele in one homologous chromosome, and the risk A-allele in the other (center panel). Thus to generate two homozygous cell lines, each harboring two copies of the risk (left panel) or protective (right panel) alleles, respectively, PE will we used in both directions simultaneously.

4.4.1. pegRNA design

Once the cleavage efficiencies of the candidate pegRNA spacers had been assessed and verified, extension sequences had to be designed for the full pegRNAs. For Spacer-01, an extension had previously been manually designed in our lab (Ext1), and while PrimeDesign had suggested the same spacer, the software opted for a slightly different extension (Ext2) by shortening the PBS with 5 nts at the 5' end, and simultaneously extending the RT template with 5 nts in the 3' direction. For Spacer-02 a single extension was chosen (Table 4.2 and Table 2.4.2).

Table 4.2. | pegRNAs designed for editing SNP rs179993

pegRNA	Spacer	Cut-to-SNP distance (nt)	Extension	Edits from → to	RT template length (nt)	PBS length (nt)	GC-content of PBS
peg1_A **peg1_C**	Spacer-01	9	Ext1_A Ext1_C	C→A A→C	14	15	60 %
peg2_A **peg2_C**	Spacer-01	9	Ext2_A Ext2_C	C→A A→C	19	10	70 %
peg3_A **peg3_C**	Spacer-02A	21	Ext3_A Ext3_C	C→A A→C	26	12	40 %

4.4.2. Cloning of pegRNA inserts into target plasmids

The pegRNAs were assembled by Golden Gate assembly and inserted into the linearized pU6-pegRNA-GG-acceptor vector backbone and transformed into competent *E.coli* bacteria. Multiple white colonies were obtained, with an estimated cloning efficiency of 30%. One clone from each reaction was picked and maxi-prepped. Following plasmid purification, the inserts were verified by Big Dye sequencing, and all were found to harbor correct sequences (Figure 4.6). Thus, all six pegRNA-plasmids were successfully generated.

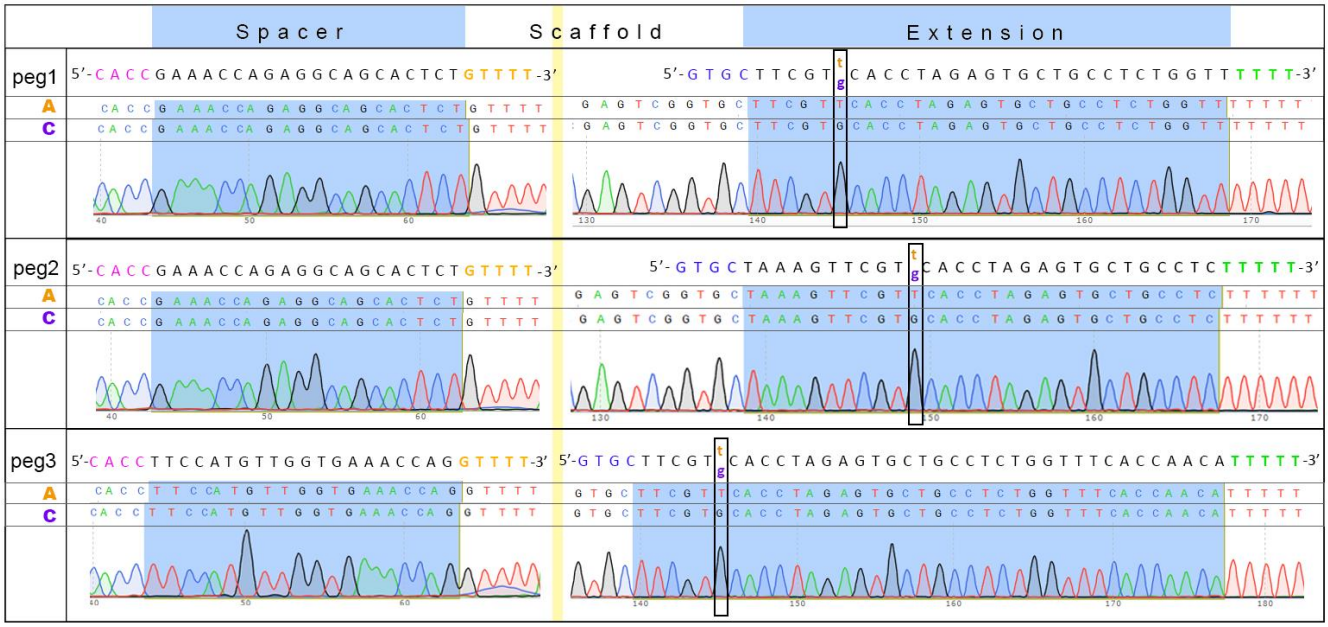


Figure 4.6 | Sequencing results of pegRNA inserts in vector pU6-pegRNA-GG-acceptor backbone. The pegRNA-encoding plasmids (containing the three pegRNAs peg1, peg2, and peg3, each with an (A) and (C) version, for a total of 6 variants) were generated by Golden Gate assembly. Each insert was assembled from 3 DNA segments; the Spacer (left, blue highlight), Scaffold (middle, white with yellow stripe (to indicate portion of sequence not shown; identical for all)) and the Extension (right, blue highlight). Expected sequence is represented above each segment, with the A/C-alleles of the extensions (for which the pegRNAs edit to) is indicated with black rectangles. The GG-overhangs are indicated with various colors flanking each segment. Each pegRNA pair (A and C-versions) has been aligned, and only one of the chromatogram displayed per pair. N = 2 sequencing replicates per insert. Chromatograms from SnapGene Viewer.

4.4.3. Optimization of transfection for Prime Editing

The PE procedure requires that two or three different plasmids are transiently transfected into the cell nucleus simultaneously; the plasmid encoding the Prime Editor-2 fusion protein (consisting of Cas9n and RT), as well as the plasmids encoding the pegRNA and ngRNA (if using the PE3/PE3b system). Since mesenchymal stem cells are notoriously difficult to transfect with typical transfection efficiencies around 10 % (Kelly et al., 2016), and considering that the pCMV-PE2-GFP plasmid is exceptionally large at 10.5 kb (which is known to negatively affect transfection efficiency (Lesueur et al., 2016)), optimizing the transfection in ASC52telo cell

line was necessary. First, however, transfection of the pCMV-PE2-GFP plasmid, followed by monitoring of GFP expression was validated in the easy-to-transfect HEK293T cell line.

Validation of the pCMV-PE2-GFP plasmid in an easily transfected cell line (HEK293T)

To ensure that the pCMV-PE2-GFP plasmid produced detectable GFP, the vector was first assessed in an easily transfected cell line, HEK293T, using the common transfection reagent TransIT-LT1 at three different DNA-to-reagent ratios (Figure 4.7). The plasmid yielded strong GFP-expression with the first GFP-positive (GFP+) cells being visible 8 hours post-transfection, and numbers increasing steadily for the duration of monitoring up until 52 hours (Figure 4.7 A).Interestingly, the lowest DNA-to reagent ratio (1:2) was found to consistently give the highest number of transformants compared to the other ratios (Figure 4.7 A). In line with this finding, the 1:2 ratio also resulted in the lowest cytotoxicity as measured by % confluency (Figure 4.7 B). Of note, compared to non-transfected (NT) or no-DNA (reagent only) controls, even the 1:2 ratio did display toxicity, but much less pronounced than for the 1:5 ratio (Figure 4.7 B).

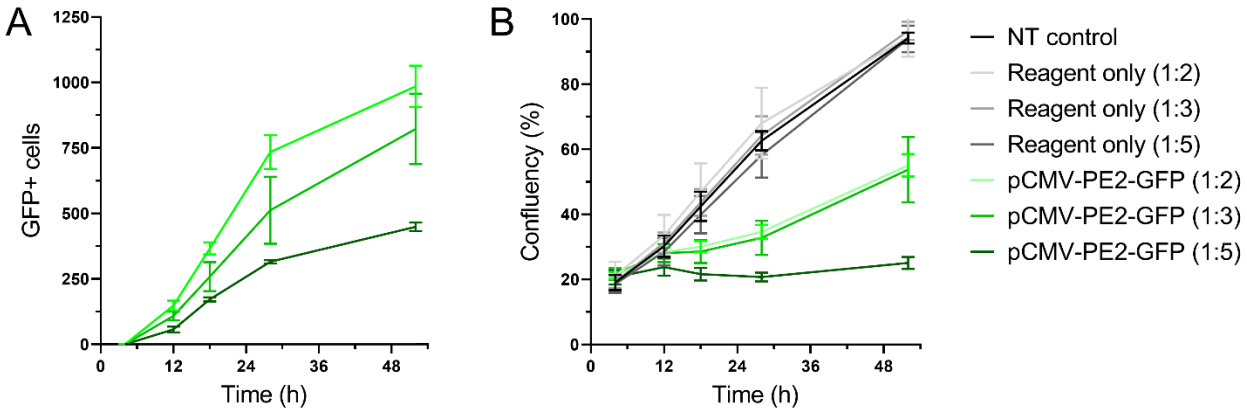


Figure 4.7 | Transfection efficiency and viability of HEK293T cells transfected with the pCMV-PE2-GFP plasmid (120 ng/7000 cells) using transfection reagent TransIT-LT1 at DNA-to-reagent ratios ranging from 1:2 to 1:5. A) Average number of GFP-positive cells per field. Each well is represented by a merge of 4 fields at 10X magnification in the Incucyte S3 system, together covering the major part of the well. N=3 replicate wells from a single experiment. Error bars represent standard deviations. **B)** Percent confluency as a measure of cell viability/proliferation. NT, non-transfected controls. PE2-GFP, pCMV-PE2-P2A-GFP.

Next, the performance of the pCMV-PE2-GFP plasmid was compared with another vector of similar size, the pDG461-Cas9n(D10A)-EGFP plasmid, which also expresses Cas9 nickase and GFP but without the reverse transcriptase. To correct for potential autofluorescence, a third plasmid lacking GFP (pU6-pegRNA-GG-acceptor-empty) was also included. The plasmids were transfected in the HEK293T cells using TransIT-LT1, as before, as well as TransIT-X2; a newer variant of the reagent which has produced higher transfection efficiencies in some cell lines in the lab (J-I. Bjune, personal communication, August 2022). As expected, the

GFP-negative vector did not result in green fluorescent cells, while the two GFP+ vectors produced high and similar levels of GFP+ cells (Figure 4.8). Interestingly, the TransIT-X2 reagent did not improve the total number of GFP+ cells (Figure 4.8 A), and was more toxic compared to the LT1 reagent (Figure 4.8 B and Figure 4.9). Thus, when dividing the number of GFP+ cells by the percent confluence per field (as a surrogate measure of transfection efficiency), the *apparent* efficiency was higher for the X2 reagent than for LT1 (Figure 4.8 C). However, because this number was mainly driven by reduced cell count, as opposed to increase in absolute number of transfected cells, LT1 was found to be a better choice overall. Taken together, both the pCMV-PE2-GFP, as well as the pDDG461-Cas9n-EGFP plasmid robustly produced a high number of GFP+ cells using different transfection reagents.

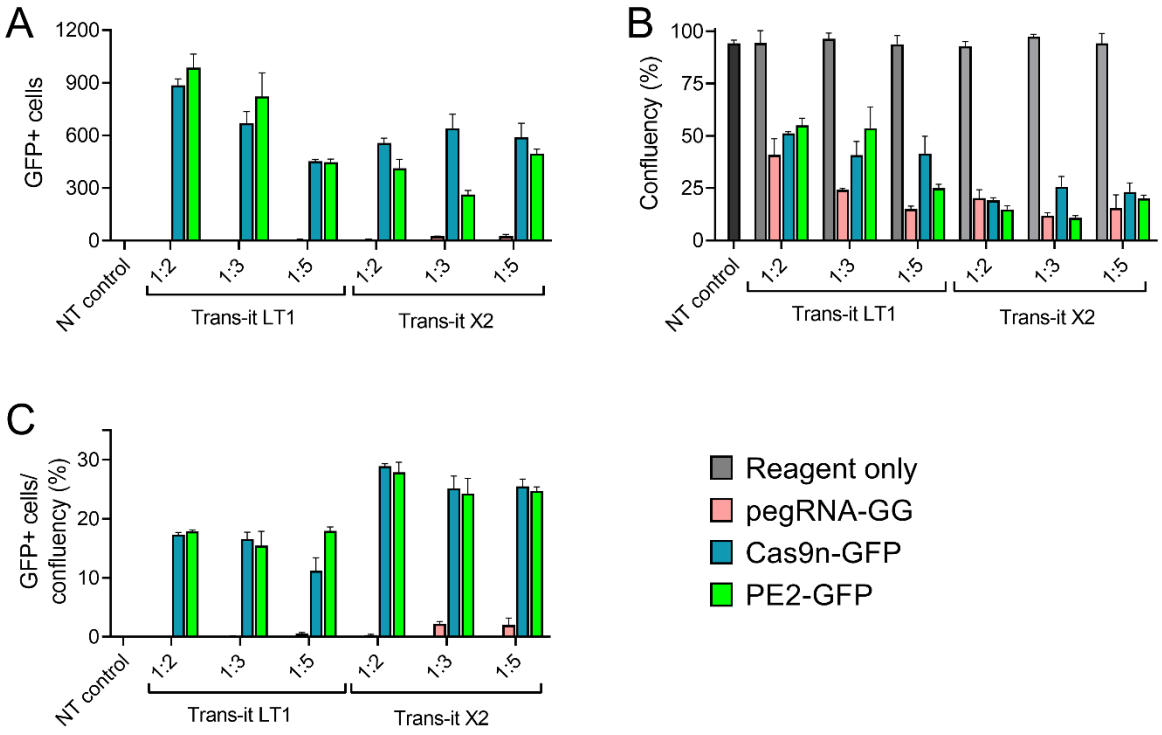


Figure 4.8 | Quantification of transfection efficiencies in HEK293T cells using two large GFP+ vectors and two different reagents (TransIT-LT1 and TransIT-X2). Each well was transfected with 120 ng plasmid and the GFP expression and confluence was measured after 52 hours. **A)** Numbers of GFP+ cells detected per field. Each well is represented by a merge of 4 fields. **B)** Percent confluency pr field as a measure of cell viability/proliferation. **C)** Surrogate transfection efficiency as measured by number of GFP+ cells divided by the percent confluence. Bars represent average values from n=3 replicate wells from a single experiment. Error bars show standard deviations. pegRNA-GG, pU6-pegRNA-GG-acceptor-empty; Cas9n-GFP, pDG461-Cas9n(D10A)-EGFP; PE2-GFP, pCMV-PE2-P2A-GFP.

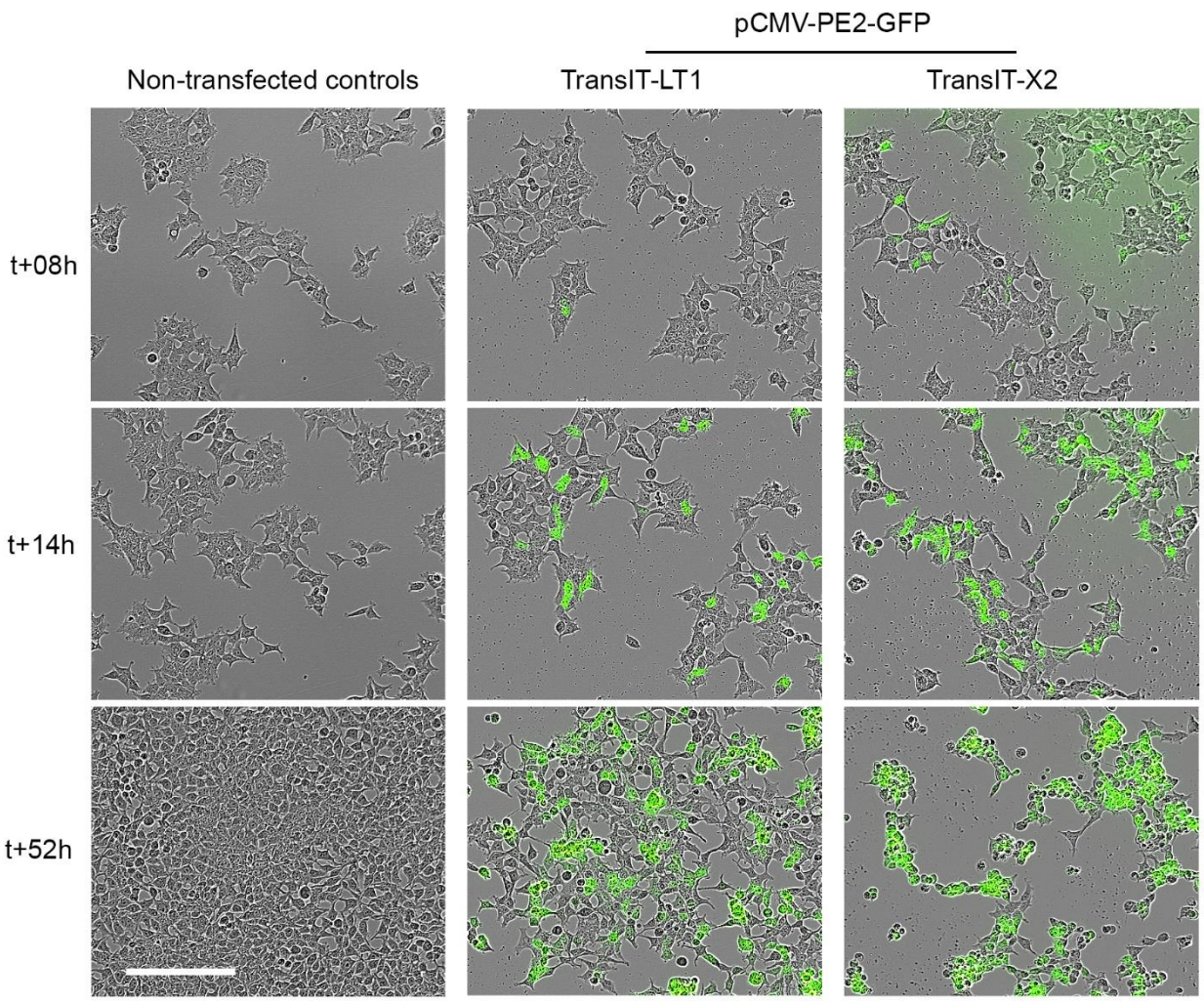


Figure 4.9 | Microscope images of GFP expression and confluence in HEK293T cells transfected with the pCMV-PE2-GFP vector. Untransfected HEK293T cells (left column), cells transfected with pCMV-PE2-P2A-GFP using reagents Trans-IT LT1 (middle column) or TransIT X2 (right column), 120 ng plasmid per well and DNA-to-reagent ratio 1:2. Each row shows the indicated number of hours post transfection (t), and each image is a merge of phase contrast and green fluorescence. A single representative image per well is shown with the same group of cells followed over time, out of N=3 wells per treatment. Scale bar = 200 μm.

Optimization of transfection in the ASC52telo cell line

Having established that the pCMV-PE2-GFP and pDG461-Cas9n-EGFP plasmids work well in HEK293T cells, the plasmids were next tested in the mesenchymal stem cell line ASC52telo (ASC52) using the same transfection reagents and conditions as before. However, since we had low stock available at the time of pCMV-PE2-GFP, and the plasmid had shown similar expression characteristics as pDG461-Cas9n-EGFP in HEK293T, the latter was chosen for initial testing in ASC52 cells.

ASC52 cells responded very differently to transfection compared to HEK293T. Very few GFP+-cells were observed, with a maximum of 8-10 GFP+ cells per field at 15 hours post transfection, coupled with extremely high levels of cell death (Figure 4.10). In fact, so much cell death occurred that using the

Incucyte software to quantify confluency was not feasible due to the presence of excessive cell debris. Moreover, after 51 hours, almost all transfected cells had died, while the non-transfected controls remained healthy (Figure 4.10, bottom row). The only difference between the two reagents was that for LT1 the cells died at a slightly slower rate than with X2.

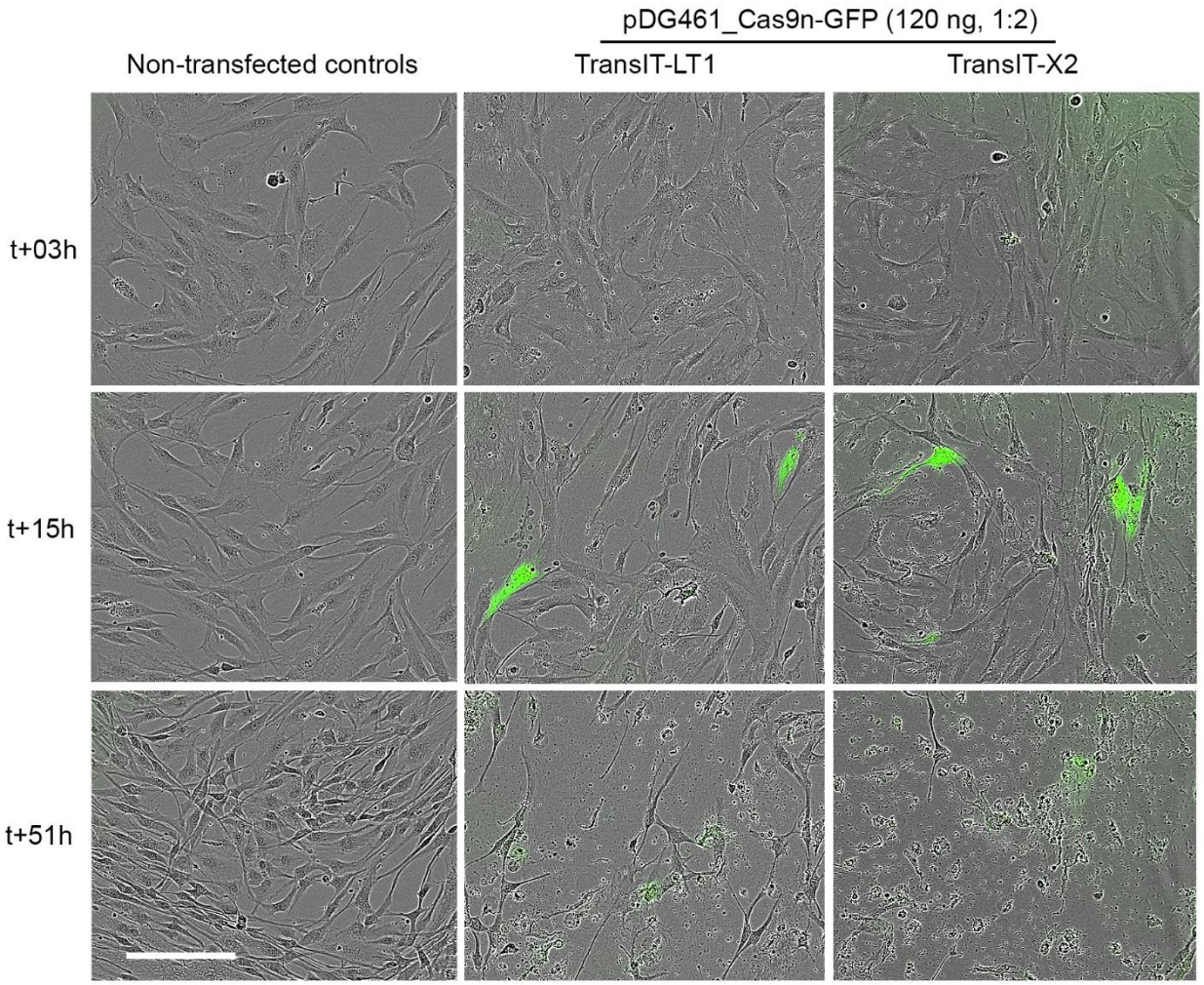


Figure 4.10 | Images of GFP expression and confluence in ASC52telo cells transfected with pDG461-Cas9n-GFP. Untransfected ASC52telo cells (left column), cells transfected with pDG461-Cas9n-GFP (120 ng/well) using reagents Trans-IT LT1 (middle column) or TransIT X2 (right column), DNA-to-reagent ratio 1:2. Each row shows the indicated number of hours post transfection (t+). For each treatment, the same group of cells is shown for the various time points. N=3 wells per treatment. Phase contrast and green fluorescence images merged. Scale bar = 200 μm.

Following this unexpectedly poor result with TransIT-LT1 and TransIT-X2, the following additional transfection reagents were tested in ASC52 in a series of experiments, to see whether any could be found that produced more viable and healthier-looking transfected cells: Xtremegene 9, Lipofectamine 2000, and Lipofectamine Stem Reagent. In addition to the pDG461-Cas9n-GFP construct, another vector of similar size

harboring the enhanced Cas9 (pCAG-eCas9-GFP-U6-gRNA) was also used, and each plasmid was tested with a range of plasmid concentrations (60, 120, and 160 ng/well) combined with three DNA-to-reagent ratios (1:2, 1:3, and 1:4) as summarized in Figure 4.11. Xtremegene 9, a polymer-based reagent, resulted in zero GFP-positive cells, though coupled with high viability (data not shown). Lipofectamine 2000, one of the leading liposomal reagents on the market, produced roughly the same number of GFP+ cells as the two TransIT reagents (Figure 4.11) and was distinctly more toxic, resulting in cell death even without the addition of plasmid DNA (data not shown). The Lipofectamine Stem Cell Reagent clearly performed better than the other reagents, resulting in 2-4 times more GFP+ cells than the other reagents (Figure 4.11), though also here high toxicity was evident.

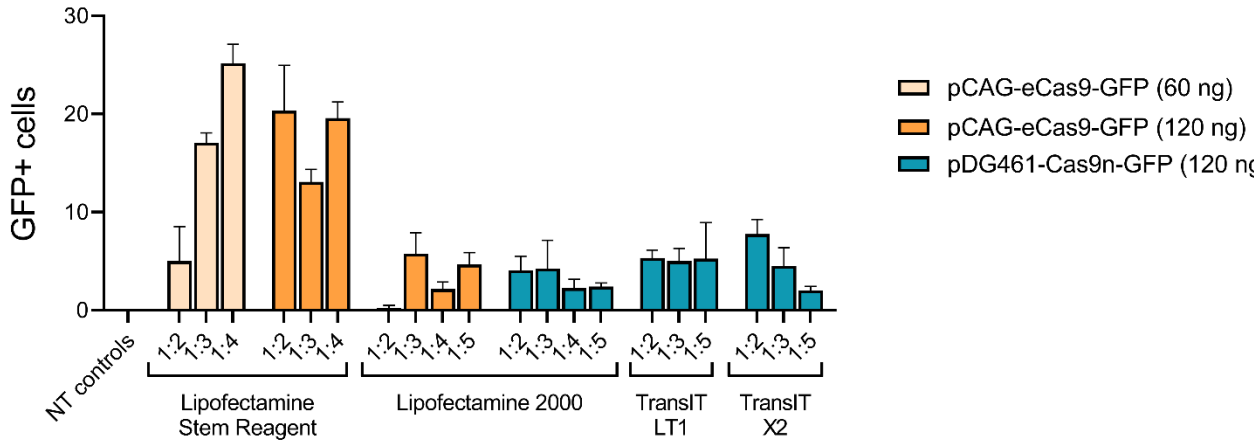


Figure 4.11 | Quantification of transfection efficiencies in ASC52telo cells transfected with one of two large Cas9-GFP plasmids using various reagents. 60 or 120 ng plasmid per well were used (as indicated in figure legend), and DNA-to-reagent ratios ranging from 1:2 to 1:5, as indicated on x-axis. Numbers of GFP+ cells were quantified per field, and each well is represented by a merge of 4 fields. For all experiments, the highest number of GFP+ cells detected at any time point was used (the optimal time point ranged from t+15-24 h). Bars represent average values from N=3 replicate wells, each from a single experiment; error bars show standard deviations. NT, non-transfected controls; pCAG-eCas9-GFP, pCAG-eCas9-P2A-GFP; pDG461-Cas9n-GFP, pDG461-SpCas9-D10A-GFP-2x-pU6; PE2-GFP.

The Lipofectamine Stem Cell reagent was also used to transfect the PE plasmid pCMV-PE2-GFP in ASC52. Surprisingly, in contrast to when transfected with pCAG-eCas9-GFP, no GFP+ cells were obtained for the prime editing plasmid, and within 48h all transfected cells had died (Figure 4.12).

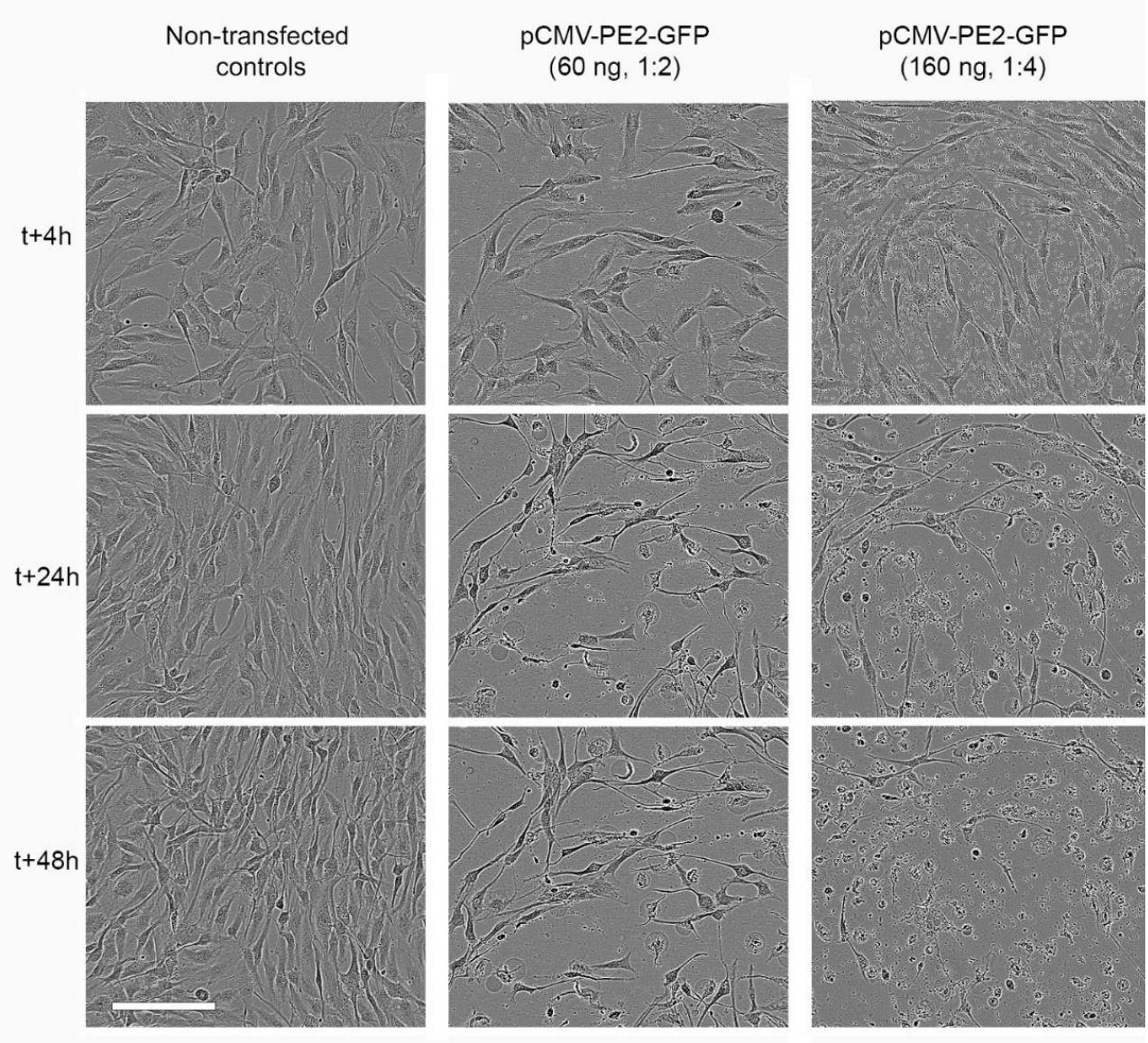


Figure 4.12 | ASC52telo cells transfected with pCMV-PE2-GFP plasmid using Lipofectamine Stem Reagent. A total of nine conditions were tested with this reagent and plasmid, all of which failed to produce a single GFP+ cell. Shown is a comparison of the mildest (middle column) and strongest (right column) conditions tested (60 ng plasmid/well, 1:2 DNA-to-reagent ratio, vs 160 ng/well and 1:4 ratio, respectively). N = 3 wells per treatment. Left column – non-transfected control cells. For each treatment, the same patch of cells is viewed for each time point. At 4 h post-transfection (t+4h), the transfection complexes are visible in the images of transfected cells, especially top right image (small specks). Medium was changed at t+5 h to remove transfection complexes, and every day following. Scale bar = 200 μm.

Because the plasmids tested thus far were large (approx. 10 kb), we next decided to test a small (3.9 kb) vector expressing only GFP (pmaxGFP) with the Lipofectamine Stem Reagent. With this plasmid-reagent combination, much higher levels of GFP+-cells were seen (Figure 4.13, pink bars), with numbers 6-10 fold higher than those achieved with pCAG-eCas9-GFP. Moreover, a small fraction of the transfected cells were still alive at 5 days following transfection (not shown). However, a strong cytotoxic effect was definitely

present, and surviving cells did not look healthy (Figure 4.14). Taken together, the PE plasmid was not compatible with the tested chemical transfection methods in the ASC52 cells.

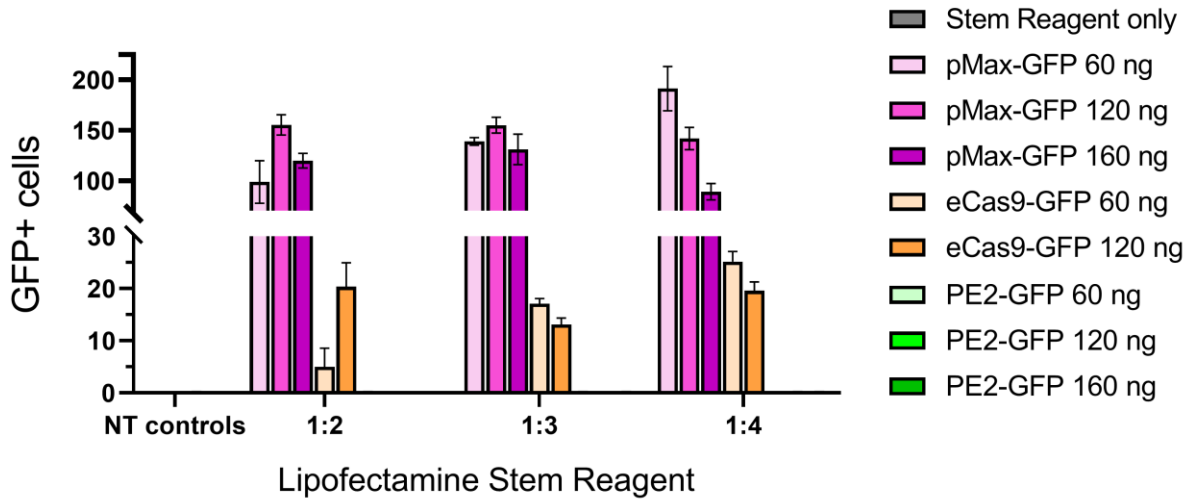


Figure 4.13 | Quantification of number of GFP+ cells produced from transfection of ASC52telo cells with 3 different GFP-expressing plasmids, together with Lipofectamine Stem Reagent. Each plasmid was tested with 2-3 different concentrations (60, 120 and 160 ng/well) and 3 different pDNA-to-reagent ratios (1:2, 1:3, 1:4). N = 3 wells per treatment. NT controls, Non-transfected controls; eCa9-GFP, pCAG-eCas9-GFP; PE2-GFP, pCMV-PE2-P2A-GFP.

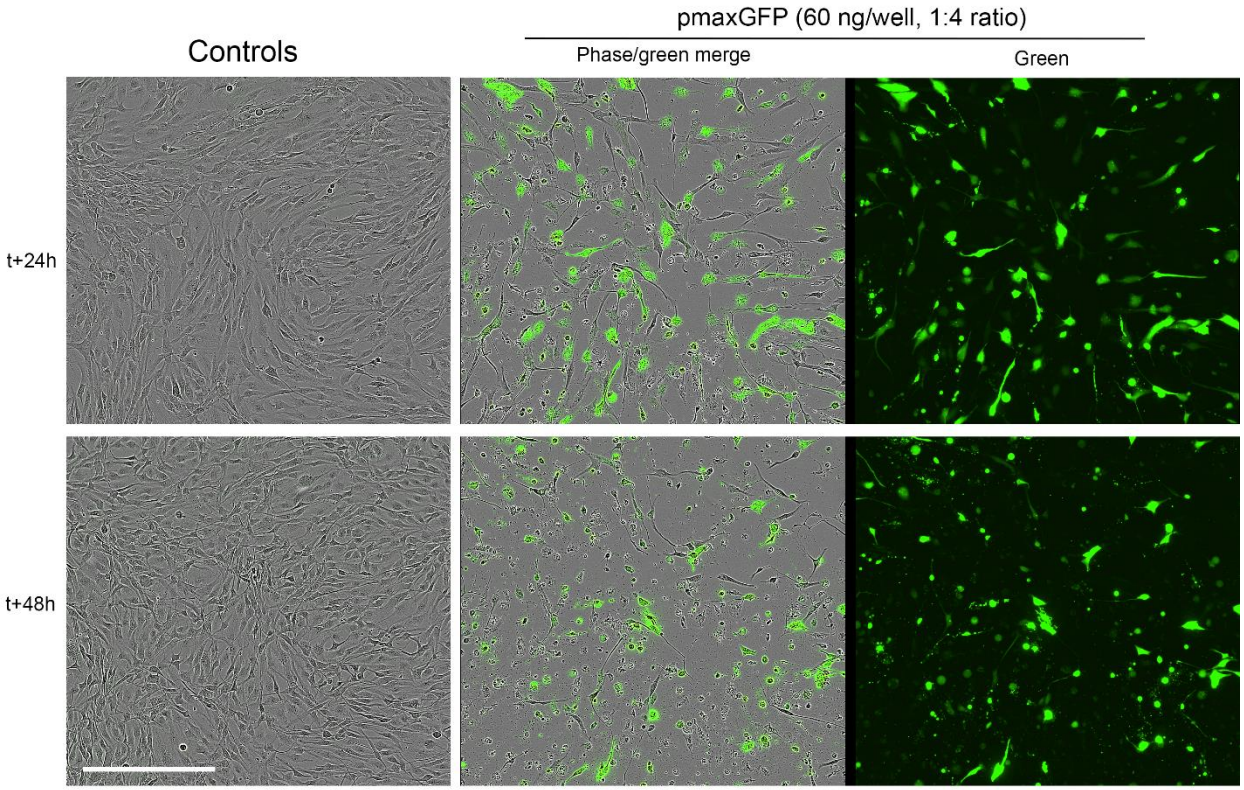


Figure 4.14 | ASC52telo cells transfected with pmaxGFP plasmid (3.9 kb) and Lipofectamine Stem reagent at 24 and 48 h post transfection. Combinations of various plasmid concentrations (60, 120 and 160 ng) and DNA-to-reagent ratios (1:2, 1:3 and 1:4) were tested. N=3 wells per treatment. Shown (two right columns) are the conditions which gave the highest transfection efficiency as assessed by number of GFP+ cells (60 ng/well pDNA, 1:4 ratio). Center column shows phase contrast/green fluorescent merge, while right column shows green fluorescent channel only. Left panel – non-transfected control cells, with same initial seeding density as the transfected cells. Scale bar = 400 μm.

Electroporation and nucleofection

Following the difficulties with chemical transfection of ASC52 cells, physical methods of plasmid-DNA delivery were attempted, including electroporation and nucleoporation. Traditional electroporation was tested with the pDG461-Cas9n-GFP plasmid at a range of concentrations and voltages (in the adipocyte progenitor cell line A41WAT), but although high cell viability was seen, there were no GFP+ cells (data not shown). Nucleofection however gave more promising results. An initial test was run with the pmaxGFP-plasmid at a range of concentrations (200-800 ng/ 50k cells) using the P1-program developed by the manufacturer for nucleofection of Primary MSCs with two different pulse conditions (EW-104, optimized for viability, and FF-104 for transfection efficiency). Although both pulse conditions resulted in the loss of a large fraction of the cells (which simply never attached following seeding), very high transfection efficiency was seen with the remainder of the cells (Figure 4.15). Unfortunately, the Incucyte was unavailable during the

days following the nucleofection test due to technical issues, and thus images were captured manually with the Nikon Eclipse Fluorescence microscope at t+24 h. Of the two pulse conditions, EW-104 (optimized for viability) resulted in lower cell loss and consequently higher numbers of transfected cells. Furthermore, no evidence of cytotoxicity was seen in the transfected cells that had survived the electric pulses, which was very encouraging.

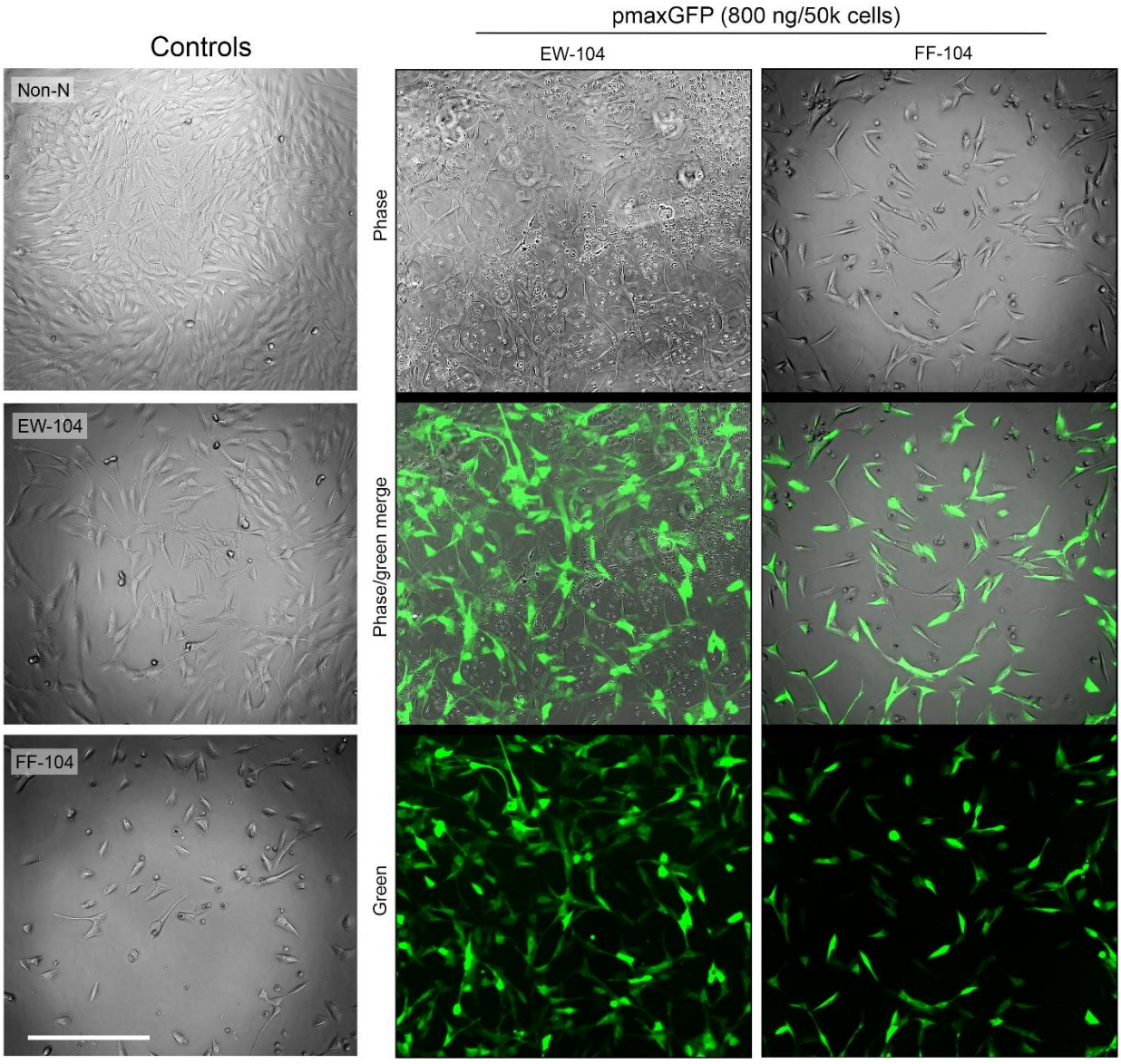


Figure 4.15 | Nucleofection-mediated transfection of ASC52 cells with plasmid pmaxGFP at 24 h post transfection. Left column: Non-transfected controls; the top was non-nucleoporated (Non-N), while the two lower images show the effects of the two pulse conditions tested; EW-104 (optimized for viability) and FF-104 (optimized for transfection efficiency). **Two right columns:** Cells transfected with 800 ng pmaxGFP per 50 000 cells**,** and pulsed with the EW-104 (middle column) or FF-104 (right column) programs. N=3 wells per treatment. For the two transfected columns (right), top image shows phase contrast, middle shows merge, and bottom shows green

fluorescence only; imaged with Nikon Eclipse Fluorescence microscope. 12 250 cells seeded/well of 96-well plate. Scale bar = 400 μm.

Next, the nucleofection conditions that yielded the best results (800 ng plasmid per 50 000 cells and EW-104 pulse condition) were used to co-transfect the three PE-plasmids in ASC52telo, with the goal of performing PE of rs1799993 in these cells. For each of the 6 pegRNA sequences designed (see Table 4.2), 50k cells were transfected with a total of 820 ng plasmid (600 ng pCMV-PE2-GFP plasmid, 120 ng pU6-pegRNA plasmid, and 60 ng ngRNA(C)). To control for the effects of co-transfection, one reaction with pCMV-PE2-GFP only (600 ng) was also run. The cells were monitored for the following two days, but disappointingly not a single GFP+ cell was observed (Figure 4.16).

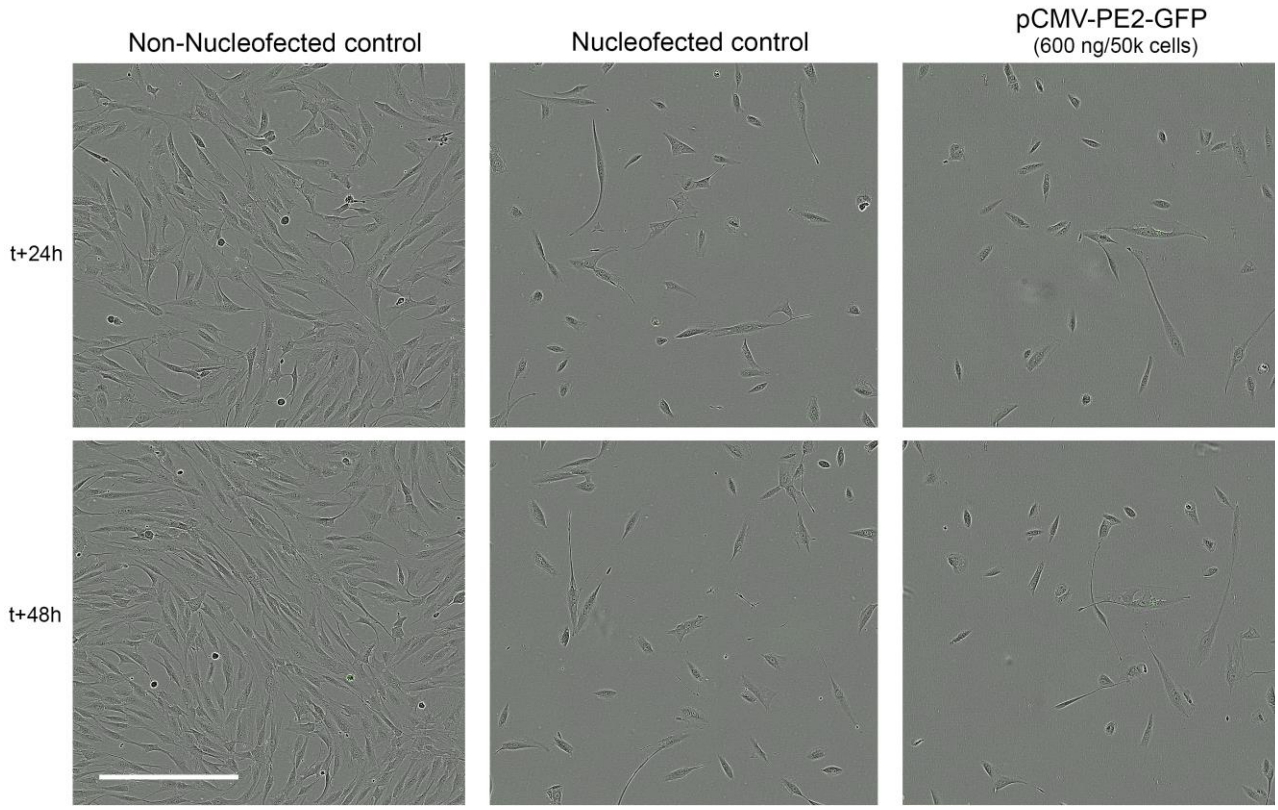


Figure 4.16 | Nucleofection of ASC52telo cells with pCMV-PE2-GFP, at t+24 h (top row) and t+48 h (bottom row). Left panel: Non-nucleofected control. Center panel: Nucleofected control, no plasmid, pulse program EW-104. Right panel: Cells transfected with 600 ng pCMV-PE2-GFP per 50 000 cells**,** and pulsed with the EW-104 program. N=1 well per treatment (of 6-well plate). Imaged with IncuCyte S3, phase contrast and green fluorescent channel merged. (Note: confluency not comparable to Figure 4.15 due to different plating formats and cell seeding density). Scale bar = 400 μm.

A confluency analysis was performed on the cells at t+24 h, with numbers of pulsed cells normalized to the non-nucleoporated control. This revealed that only 17 % of cells survived the pulsing when no plasmid was added (Figure 4.17 A, NC – nucleoporated control). Furthermore, when plasmids were added, survival

rate decreased to <10 % (Figure 4.17 A, colored bars). Whether this additional loss occurred during pulsing or subsequently is unknown, but when the growth rate was analyzed for the following 24 h (from t + 24 h to t + 48 h), all plasmid-nucleoporated cells had further decreased in confluency, in contrast to the nucleoporated control, which had an incremental positive growth rate of ~3 % (Figure 4.17 B). For comparison, non-nucleoporated controls had a growth rate of 30 % in the same time span.

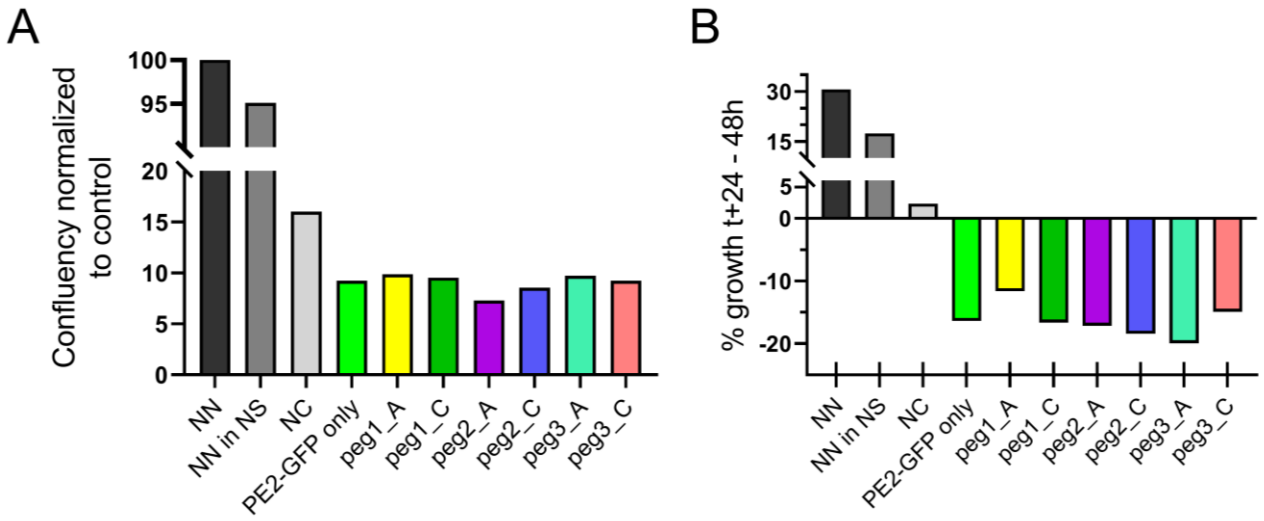


Figure 4.17 | Effect of Nucleofection of PE-plasmids on ASC52telo using pulse conditions EW-104. A) Confluency of nucleofected cells relative to non-nucleofected (NN) control at 24 h post-pulsing. **B)** Relative change in confluency from t+24 h to t+48 h. NN in NS – Non-nucleofected cells incubated in nucleofection solution. NC – Nucleofected, plasmid-free controls. PE2-GFP only – cells nucleofected with 600 ng pCMV-PE2-P2A-GFP per 50 000 cells. peg1_(A or C) to peg3_(A or C) – Cells nucleofected with a total of 800 ng plasmid (600 ng pCMV-PE2-GFP, 120 ng pU6-pegRNA(number_A or C) plasmid, and 60 ng ngRNA(C)) per 50 000 cells.

Overall, the nucleofection transfection method was found to work perfectly with the small pmaxGFP plasmid in the ASC52cells, but even this method was not able to bring the much larger PE plasmid into the cell nucleus for successful expression. Further optimization was not performed due to time limitations.

4.5. CRISPR/Cas9-mediated enhancer modulation

Following the poor success of transfecting ASC52telo cells for PE, we decided to switch our focus to CA/I. For this method, stably transfected cell lines of ASC52 would be generated by lentiviral transduction, for inducible expression of the effector-dCas9-fusion proteins, and constitutive expression of the sgRNAs and MCP-TF proteins. Since lentiviral transduction in many cases has been shown to be more successful than transfection in mesenchymal stem cells (Zhou et al., 2023), this method would hopefully be more successful. Both CA and CI would be performed in parallel, so that the downstream effects of the enhancer could be detected by comparing gene expression results from cells where the enhancer was either fully active or fully repressed.

4.5.1. sgRNA plasmid assembly by Golden Gate Cloning

Based on the *in vitro* cleavage assay of the sgRNA spacer sequences (Section 4.2), a variant of each of the 5 spacers was selected for use with CA/I; of each A and B version, the one with the best cleavage efficiency was selected. If both versions showed equal efficiency, the B-version (starting with a G) was used, since having a G as the first nucleotide is predicted to increase sgRNA transcription from the U6 promoter.

For CA/I, Lentiviral vectors containing the sgRNA scaffold including the specially designed MS2-hairpins were used as the plasmid backbones, for which the spacer sequences were inserted. Downstream of the sgRNA sequence, the vectors carried a second open reading frame containing the MCP-TF fusion protein (VP64 or KRAB for activation or repression, respectively), and so 10 plasmids had to be generated; one set each containing all 5 spacers for both CA and CI. This was achieved by GG-assembly, and the amplified plasmids were checked for correct insert by BigDye sequencing. Nine of the 10 plasmids were successfully generated (Figure 4.18).

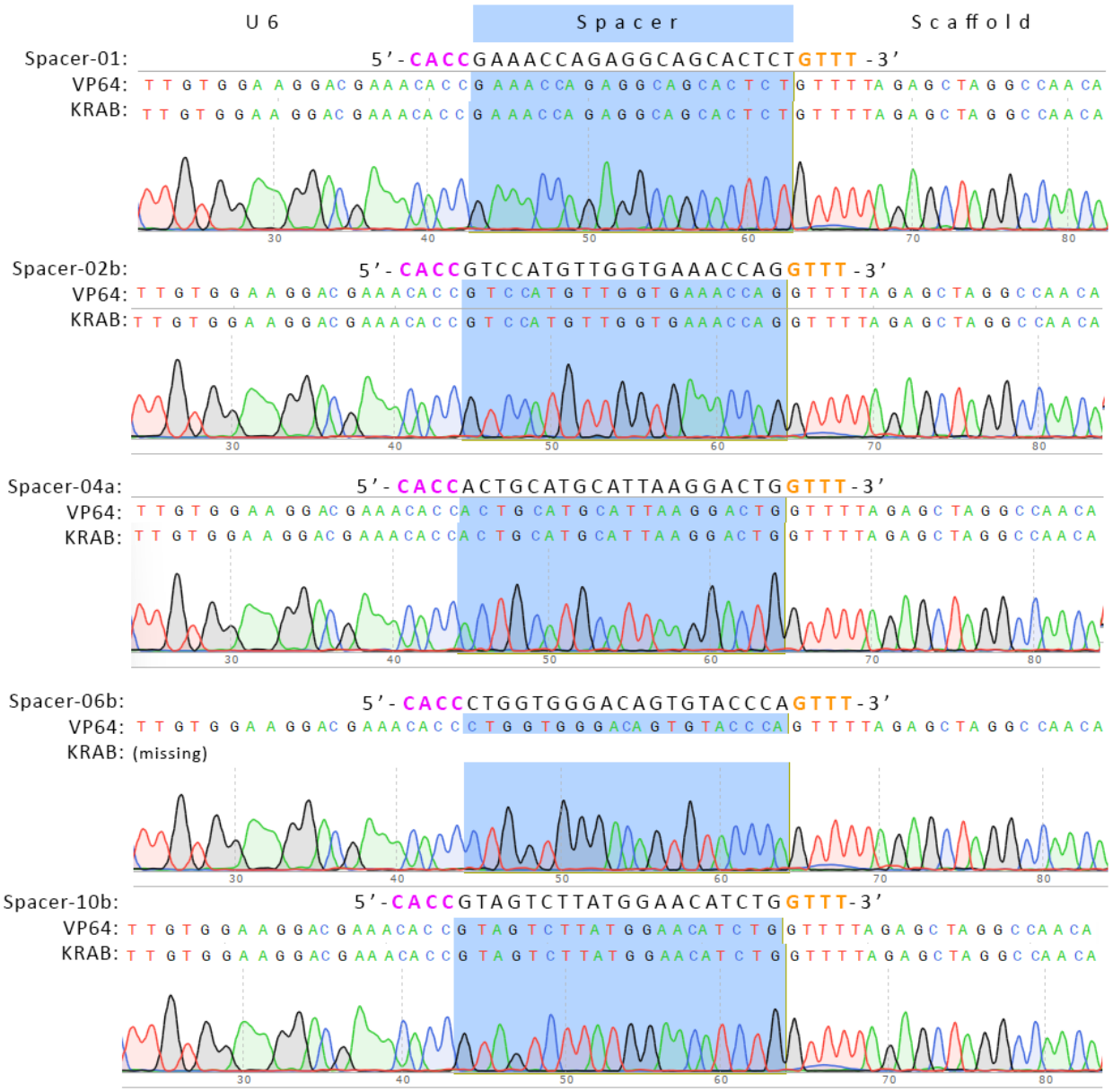


Figure 4.18 | Sequencing results of inserts in pLenti-sgRNA plasmids. Spacers were inserted into the pLenti_sgRNA(MS2)_MCP-TF-zsGreen1 plasmids (TF = VP64 or KRAB, for CA or CI, respectively) by GG-assembly and transformed into competent *E.coli* for amplification. Inserts were sequenced (blue highlights) and verified against the expected sequence (indicated above each sequencing result, including the flanking GG-sequences used for assembly). Each spacer was inserted into both the VP64 (CA) and the KRAB (CI) editions of the sgRNA expression plasmid, and thus the sequencing results from both are aligned, and only one chromatogram displayed per pair. N = 2 sequencing reactions per plasmid.

4.5.2. Lentiviral production in packaging cell line HEK293T

To produce the lentiviral vectors for transduction of ASC52telo, each transfer plasmid was co-transfected with the envelope and packaging plasmids into HEK293T cells. The expression cassette from each plasmid would be packaged as a dsRNA lentiviral vector genome for integration into the host genome. A total of 12 lentiviral vector constructs were prepared; one each for the plasmids Tet-on-3G-BFP (Figure 4.19 A), the two effector-dCas9 plasmids with inducible TRE3G promoters (p300-dCas9-P2A-mCherry and LSD1-dCas9-P2A-mCherry) (Figure 4.19 B), and the 9 sgRNA plasmids; pU6-sgRNA-*Spacer*(MS2)_EF1a-MCP-*TF*-IRES-zsGreen1, where spacer = the various spacer sequences (Tables 4.1 and 2.4.1), and TF = VP64 or KRAB for CA or CI, respectively (Figure 4.19 C).

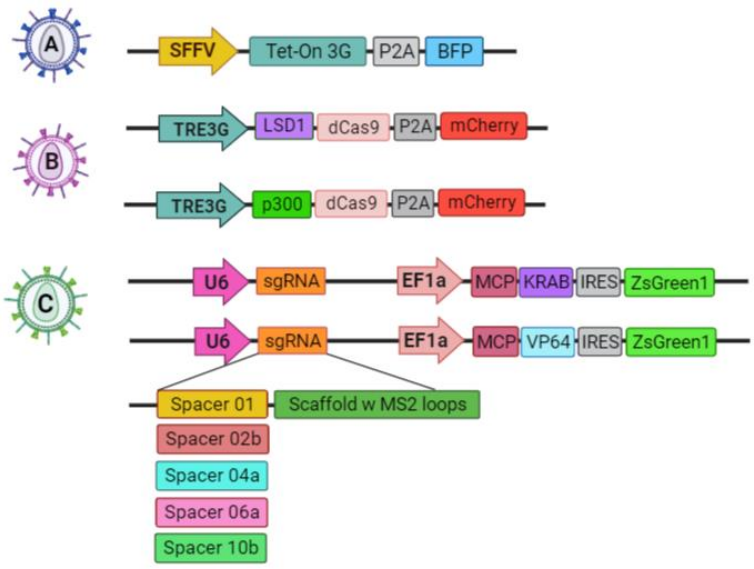


Figure 4.19 | Overview over expression cassettes incorporated into the three categories of lentivirus vectors produced in this thesis. A) A single variant of vector A was produced, containing the Tet-on-3G for inducible activation of the TRE3G promoter. **B)** Two variants of vector B were prepared, one each containing the LSD1-dCas9 and p300-dCas9 sequences for CI and CA, respectively (inducible by vector A). **C)** Nine variants of vector C (containing the sgRNAs and downstream MCP-TFs) were prepared; one set of 4 for CI (containing MCP-KRAB; variant with spacer-06a not produced) and a full set of 5 for CA (containing MCP-VP64). *Figure prepared in Biorender.*

4.5.3. Lentiviral transduction of ASC52

Two consecutive rounds of transduction were to be performed; the first was a co-transduction of Tet-on-3G-BFP and effector-dCas9-mCherry lentiviral vectors, and, following FACS sorting to select only dually transfected cells, a second round of transduction would insert one or more sgRNA(MS2) and MCP-TF sequences. The dually transduced cells would be selected based on mCherry expression, which requires both transfer sequences for expression, since the promoter of mCherry, TRE3G, requires Tet-on-3G for activation in the presence of doxycycline (which was added to the culture medium during and following transduction; see Figure 3.6, methods).

Directly before the first round of transduction, the titer of the stocks was determined by using an instant lentiviral titer test (LentiX GoStix Plus) to quantify the concentration of capsid protein p24 in each purified lentiviral vector preparation. p24 concentration ranged from 264 -714 ng/mL (Figure 4.20 and Table 4.3).

These values were then converted to number of physical viral particles (PP), and further to theoretical value for infectious units (IFU)/mL (Table 4.3). This was then used to determine volumes of lentiviral vector preparation required for the various MOIs (number of IFUs per target cell). Since dual transduction was required, but number of integration events needed to be kept as low as possible (preferably only one event for each of the two transfer sequences), a range of MOIs was tested and the lowest MOI that produced a sufficient number of doubly infected cells would be expanded and used for the next transduction. The cells were therefore infected with MOIs = 1, 5, 10 and 30.

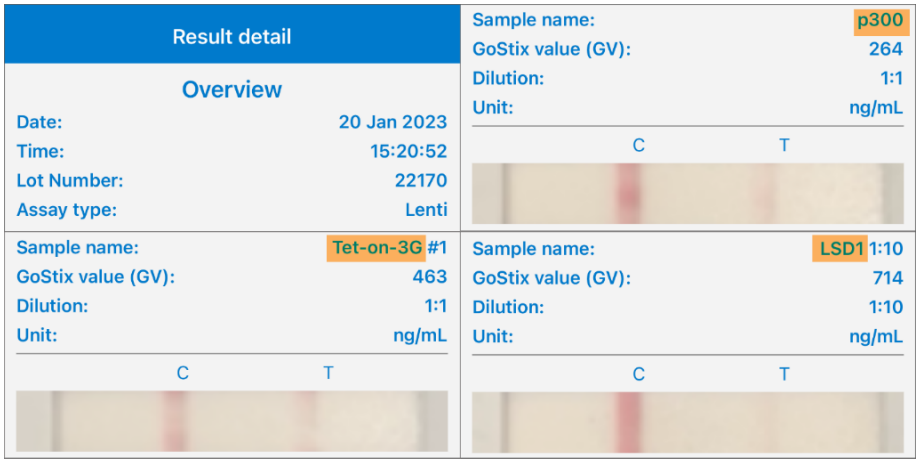


Figure 4.20 | Quantification of a proxy for lentiviral titer, using LentiX GoStix. To estimate viral titer, the concentration of viral capsid protein p24 was measured in the concentrated preparation by the use of LentiX GoStix plus. A small sample of vector preparation was applied to a test cassette, and the strength of the test band indicated concentration (line at T; C = control line). The strength of the band was quantified by an accompanying app, which returned a CV value equivalent to p24 (ng/ml). Sample name highlighted in yellow.

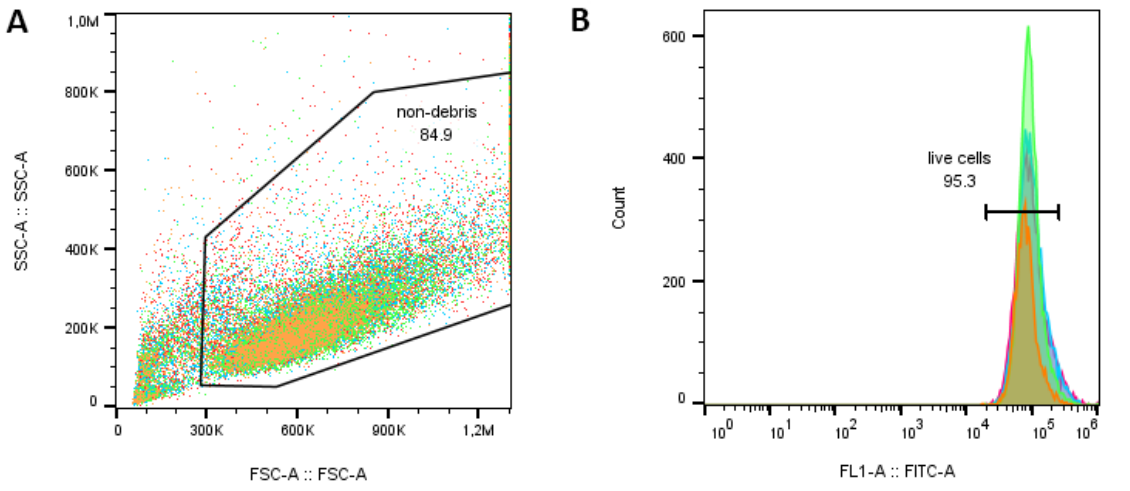
Table 4.3 | Quantification of viral titer for vectors used in Transduction #1.

Virus	GV = p24(ng/mL)	PP/mL	IFU/mL
Tet-on-3G (48h)	**463**	$4.63*10^9$	$5.79*10^7$
p300-dCas9 (48h)	**264**	$2.64*10^9$	$3.30*10^7$
LSD1-dCas9 (48h)	**714**	$7.14*10^9$	$8.93*10^7$

GV, LentiX GoStix value; PP, physical viral particles; IFU, infectious units.

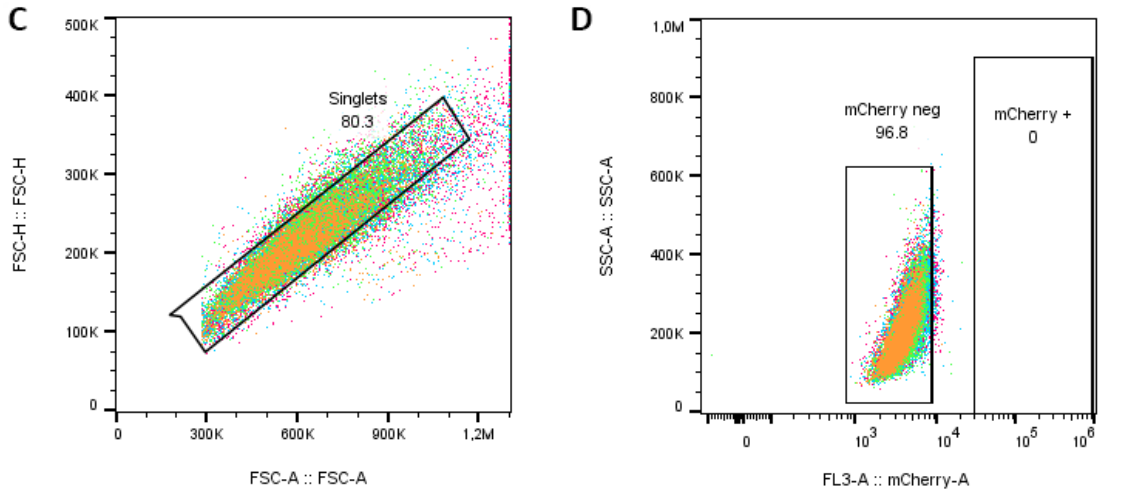
4.5.4. Flow cytometry and fluorescence-assisted cell sorting of transduced cells

Three days following the first transduction of ASC52, the cells were harvested and transduction efficiencies determined by flow cytometry. A Live/Dead fluorescent marker was used with the CA-transduced cells to exclude dead cells, which indicated that all cells were alive, and was therefore eliminated for the CI-transduced cells. A series of 4 gates was set up based on non-transduced controls, for the following: 1) Removal of non-cell debris, 2) elimination of dead cells with high live/dead marker fluorescence (this gate was left open for samples without the marker), 3) elimination of non-single cells, and 4) detection of mCherry+ cells. However, when the transduced cells were analyzed, for both CA and CI with all MOIs, zero mCherry+ cells could be detected (Figure 4.21 and Table 4.4).



	Sample Name	Subset Name	Count
	a2 MOI=1	Ungated	5578
	a3 MOI=5	Ungated	10000
	a5 MOI=30	Ungated	10000
	a6 MOI=30 NO DOX	Ungated	10000

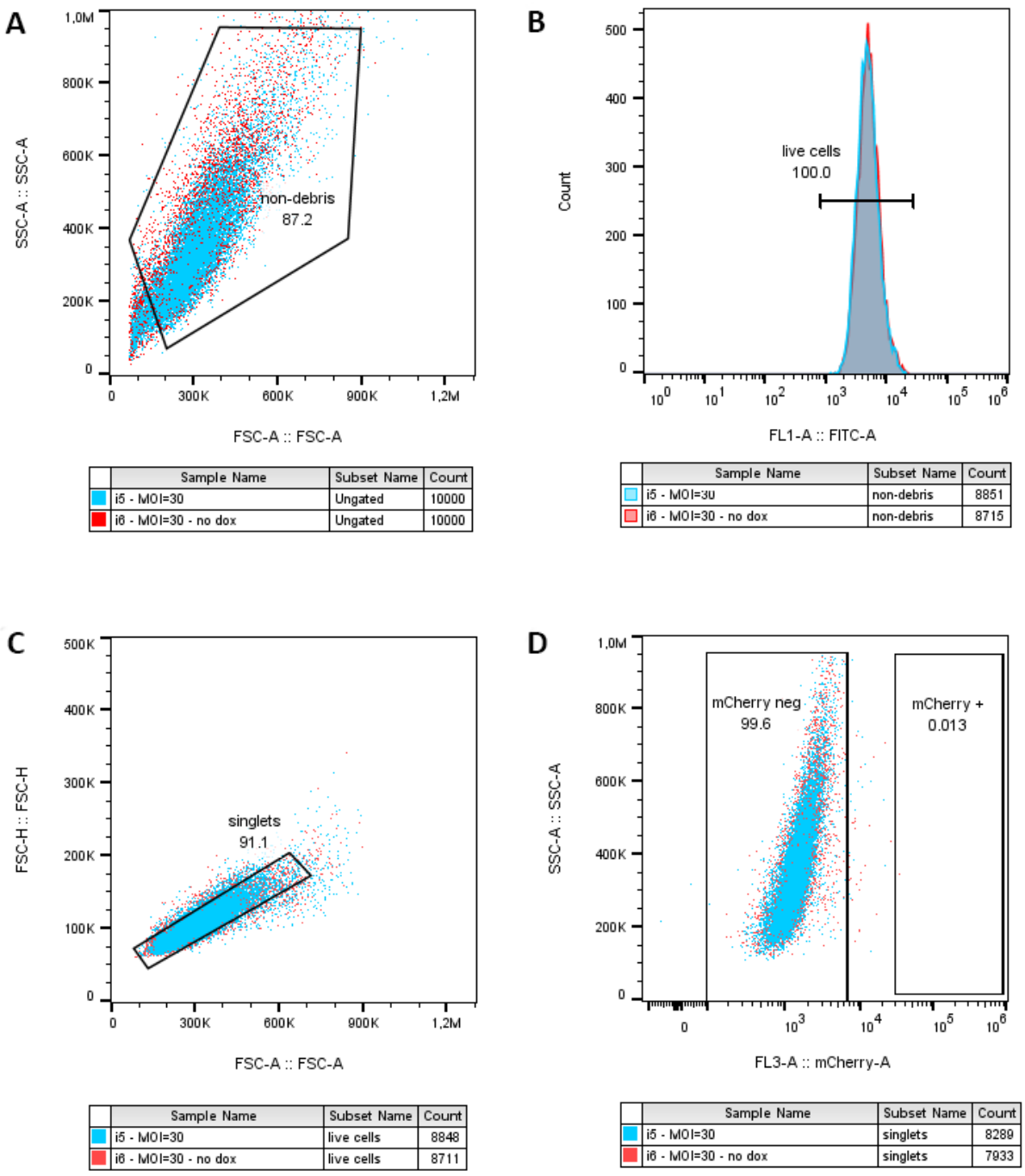
	Sample Name	Subset Name	Count
	a2 MOI=1	non-debris	4878
	a3 MOI=5	non-debris	9062
	a5 MOI=30	non-debris	8648
	a6 MOI=30 NO DOX	non-debris	8492



	Sample Name	Subset Name	Count
	a2 MOI=1	live cells	4422
	a3 MOI=5	live cells	8227
	a5 MOI=30	live cells	7959
	a6 MOI=30 NO DOX	live cells	8097

	Sample Name	Subset Name	Count
	a2 MOI=1	Singlets	4166
	a3 MOI=5	Singlets	7671
	a5 MOI=30	Singlets	7221
	a6 MOI=30 NO DOX	Singlets	6503

Figure 4.21 | Flow cytometry analysis of mCherry in transduced cells (CA). Cells transduced with the lentiviral vectors Tet-on-3G-BFP and p300-dCas9-mCherry were analyzed for integration of both sequences by red fluorescence (mCherry expression). Samples **a2** (MOI=1, orange scatter), **a3** (MOI=5, green scatter), and **a5** (blue scatter): cells transduced with equivalent MOI of both vectors, and cultured in doxycycline for 3 days before analysis to induce p300-dCas9-mCherry expression via the inducible promoter TRE3G in cells that had been successfully transduced with both vectors. Sample **a6** (pink scatter): cells also transduced with MOI = 30 of both vectors, but not cultured in doxycycline, and were therefore not expected to express mCherry even in doubly transduced cells (control). LIVE/DEAD fixable green was included as a live/dead marker. The following gates were set up to detect successfully transduced cells: **A)** forward scatter vs side scatter (FSC-A vs SSC-A) to remove debris, **B)** forward scatter through the FL1-A filter to eliminate dead cells, **C)** area vs height (FSC-A vs FSC-H), to eliminate doublets, and **D)** forward vs side scatter through the FL3 filter to detect mCherry+ cells. Gates set up according to non-transduced cells. All evaluated samples superimposed for each gating strategy.



	Sample Name	Subset Name	Count
	i5 - MOI=30	Ungated	10000
	i6 - MOI=30 - no dox	Ungated	10000

	Sample Name	Subset Name	Count
	i5 - MOI=30	non-debris	8851
	i6 - MOI=30 - no dox	non-debris	8715

	Sample Name	Subset Name	Count
	i5 - MOI=30	live cells	8848
	i6 - MOI=30 - no dox	live cells	8711

	Sample Name	Subset Name	Count
	i5 - MOI=30	singlets	8289
	i6 - MOI=30 - no dox	singlets	7933

Figure 4.22 | Flow cytometry analysis of mCherry expression in transduced cells (CI). Cells transduced with the lentiviral vectors Tet-on-3G-BFP and LSD1-dCas9-mCherry were analyzed for integration of both sequences by red fluorescence (mCherry). Sample **i5** (blue scatter): cells transduced with MOI = 30 of both vectors, and cultured in doxycycline for 3 days before analysis, to induce LSD-dCas9-mCherry expression via the inducible promoter TRE3G in cells that had been successfully transduced with both vectors. Sample **i6** (red scatter): cells also transduced with MOI = 30 of both vectors, but not cultured in doxycycline and were therefore not expected to express mCherry even in doubly transduced cells (control). LIVE/DEAD fixable green was omitted for these samples. The following gates were set up to detect successfully transduced cells: **A)** forward scatter vs side scatter (FSC-A vs SSC-A) to remove debris, **B)** open, **C)** area vs height (FSC-A vs FSC-H) to eliminate doublets, and finally **D)** forward vs side scatter through the FL3 filter to detect mCherry+ cells. Gates set up according to non-

transduced cells. Due to lack of positive results for the highest MOIs, the cells transduced with lower MOIs were not evaluated. The two evaluated samples superimposed for each gating strategy.

Table 4.4 | Flow cytometry results of transduced cells, frequency of mCherry+ cells.

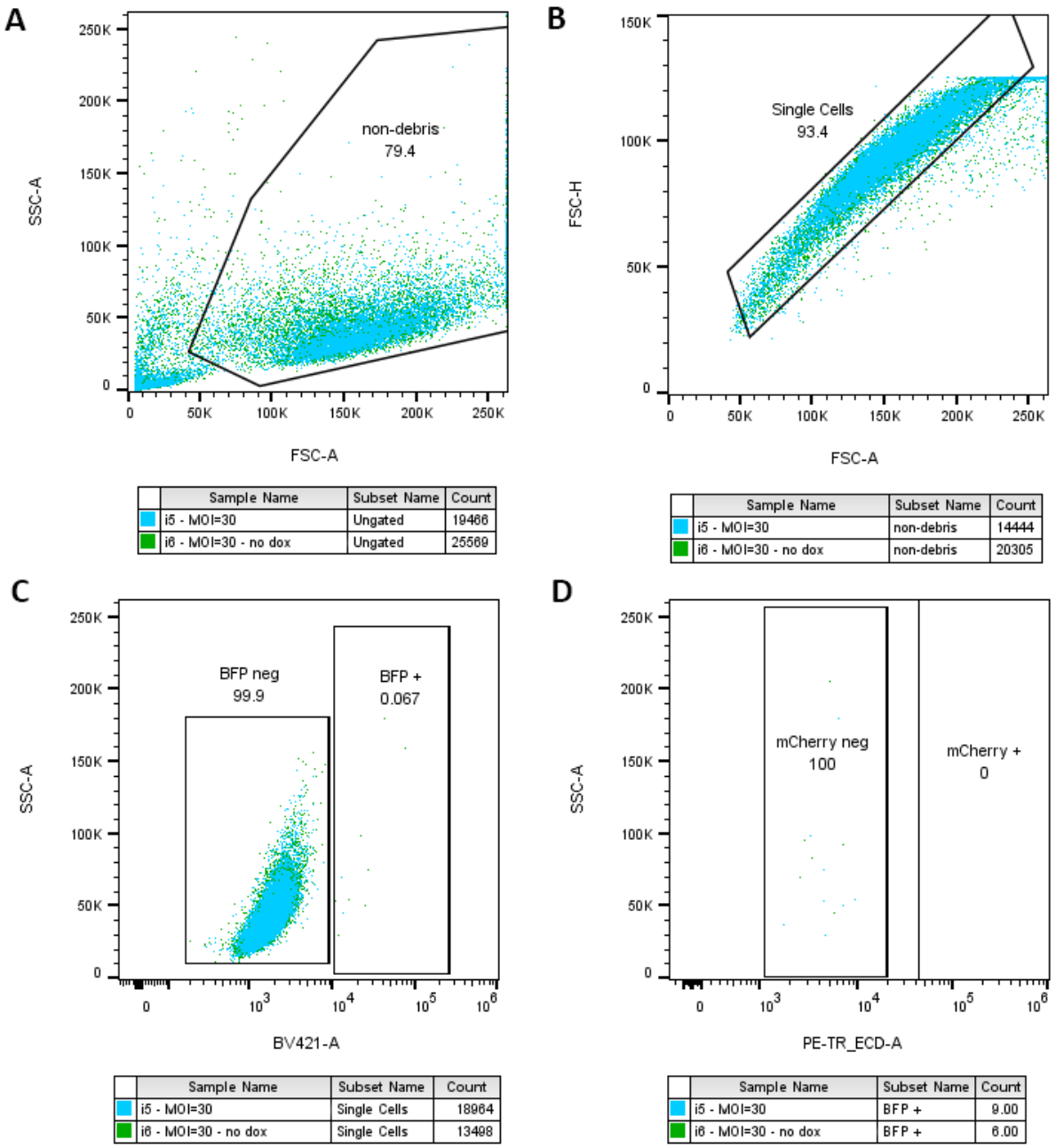
Sample	Freq. of Parent	**Non-debris** (Freq. of Parent)	**Live cells** (Freq. of Parent)	**Singlets** (Freq. of Parent)	**mCherry neg** (Freq. of Parent**)**	**mCherry +** (Freq. of Parent)
a2 MOI=1	100 %	87.5 %	90.7 %	94.2 %	99.9 %	0.0 %
a3 MOI=5	100 %	90.6 %	90.8 %	93.2 %	99.8 %	0.0 %
a5 MOI=30	100 %	86.5 %	92.0 %	90.7 %	97.7 %	0.0 %
a6 MOI=30 no dox	100 %	84.9 %	95.3 %	80.3 %	96.8 %	0.0 %
i5 - MOI=30	100 %	88.5 %	-	93.7 %	99.7 %	0.0 %
i6 - MOI=30 no dox	100 %	87.2 %	-	91.1 %	99.6 %	0.0 %

Because the expression of mCherry depended on the expression of Tet-on-3G, the lack of mCherry expression did not reveal whether the problem was with the effector-dCas9-mCherry lentiviral construct, the Tet-On-3G-BFP construct, or both. Therefore, a sample of the cells transduced with the highest MOI (=30) was analyzed for single transduction of the Tet-on-3G-BFP lentiviral vector based on blue fluorescence, using a compatible flow cytometer. Virtually no BFP+ cells were present (Figure 4.23 and Table 4.5), suggesting suboptimal lentiviral transduction of at least the Tet-on-3G vector.

In conclusion, the pilot lentiviral-mediated transduction in ASC52telo cells was unsuccessful, and the workflow was therefore terminated at this point. The poor results could have several explanations and would require extensive troubleshooting and optimization, including assessing both the packaging and delivery of the lentiviral particles. Due to time limitations, these attempts at optimization were not performed at the present time.

Table 4.5 | Flow cytometry results of transduced cells, frequency of BFP+ cells.

Sample	Freq. of Parent	**Non-debris** (Freq. of Parent)	**Singlets** (Freq. of Parent)	**BFP neg** (Freq. of Parent)	**BFP +** (Freq. of Parent)
i5 MOI=30	100 %	74.2 %	93.5 %	99.9 %	0.067 %
i6 MOI=30 (no dox)	100 %	79.4 %	93.4 %	100 %	0.032 %



Sample Name	Subset Name	Count
i5 - MOI=30	Ungated	19466
i6 - MOI=30 - no dox	Ungated	25569

Sample Name	Subset Name	Count
i5 - MOI=30	non-debris	14444
i6 - MOI=30 - no dox	non-debris	20305

Sample Name	Subset Name	Count
i5 - MOI=30	Single Cells	18964
i6 - MOI=30 - no dox	Single Cells	13498

Sample Name	Subset Name	Count
i5 - MOI=30	BFP +	9.00
i6 - MOI=30 - no dox	BFP +	6.00

Figure 4.23 | Flow cytometry analysis of BFP in transduced cells. Cells transduced with the lentiviral vectors Tet-on-3G-BFP and LSD1-dCas9-mCherry were analyzed for integration of the Tet-on-3G sequence by blue fluorescence. i5 (blue scatter), cells transduced with MOI = 30 of both vectors, and cultured in doxycycline for 3 days before analysis, to induce LSD-dCas9-mCherry expression via the inducible promoter TRE3G in cells that had been successfully transduced with both vectors. i6 (green scatter), cells also transduced with MOI = 30 of both vectors, but not cultured in doxycycline, and were therefore not expected to express mCherry even in doubly transduced cells. The following gates were set up to detect BFP-expressing cells: **A)** forward scatter vs side scatter (FSC-A vs SSC-A) to remove debris, **B)** area vs height (FSC-A vs FSC-H) to eliminate doublets, **C)** forward vs side scatter through the BV421-A filter to detect BFP+ cells, and finally **D)** forward vs side scatter through the PE-TR_ECD-A filter to detect mCherry+ cells. Number of cells detected for each gate indicated below each plot. Both samples superimposed for each gating strategy.

5. Discussion

Making biological sense out of GWAS data has long been a major limitation in understanding the connection between genetic variation and disease risk (Claussnitzer et al., 2020). In this thesis, two CRISPR/Cas9-based techniques have been adapted and tested to provide a methodological framework that addresses two of the main limitations of GWAS; identification of causal SNPs in non-coding disease-associated haploblocks, and the mapping of downstream target genes and biological pathways. As a proof-of-principle, these methods were applied to a newly discovered locus (11q23.3) associated with visceral obesity, which is a particularly harmful type of obesity with completely unknown genetic etiology (Karlsson et al., 2019; Saad et al., 2022).

Taking advantage of recent advances in functional dissection of GWAS data (Claussnitzer et al., 2015; Sinnott-Armstrong et al., 2021), this thesis builds upon previous unpublished work conducted by previous master students in our group, which identified, rs1799993 as a likely causal SNP in an enhancer of the 11q23.3 haploblock through *in vitro* mutagenesis and luciferase reporter assays (Mirza, 2022; Samuelsen, 2021). Moreover, epigenetic data suggest that the variant affects an enhancer element that is active in adipose-derived mesenchymal stem cells (AD-MSCs). To experimentally establish the causal link between the SNP and visceral obesity, prime editing of rs1799993, as well as CRISPR/Cas9-mediated activation/inhibition of the surrounding enhancer was tested in the in the AD-MSCs cell line ASC52telo. While a library of necessary plasmids for the two methods were successfully cloned, the transfection and transduction of the ASC52telo cells proved difficult, as discussed in the following sections.

5.1. Methodological considerations

5.1.1. sgRNA design

A critical component of both prime editing and CRISPR activation is well-designed guide RNAs with high target specificity and cleavage efficiency (Haeussler et al., 2016). Thus, a great effort was put into careful design of guides/spacers which resulted in excellent theoretical properties. However, because actual performance may differ from predicted (Konstantakos et al., 2022), the designed guides were functionally screened *in vitro.* Notably, all sgRNAs induced cleavage of the target DNA template at the expected locations and most of the reactions were highly efficient (Figure 4.3). A secondary goal of the assay was to assess whether there was a difference in cleavage efficiencies between the (A) and (B) versions of each sequence. Of the four spacers that did not naturally start with a G, three showed no difference in cleavage efficiency when first nucleotide was exchanged to a G, indicating that a 1 bp mismatch at the 5’ end of the spacer did not have an affect on target binding. This observation agrees with the literature, as WT Cas9 is known not to be sensitive to mismatches at the 5’ end, which is the opposite end of the cleavage site (Wang et al., 2019).

However, with Spacer-06, the (B)-version performed considerably worse than the (A)-version (Figure 4.3). The spacer sequence can affect on-target cleavage efficiency in a myriad of ways apart from

target complementarity, including destabilizing the sgRNA structure, inducing secondary structures that reduce accessibility of the spacer to the template, and spacer GC content (Konstantakos et al., 2022). Thus, the presence of a G as the first nucleotide in Spacer-06 may have resulted in unfavorable interactions for sgRNA function and Cas9 loading. Similar factors may also have contributed to the significantly lower cleavage efficiency observed for both versions of Spacer-04.

The fact that the ngRNA spacer sequences did not display the hypothesized genotype-specificity was not entirely unexpected, considering that the selective nucleotide of the ngRNA was at the 3rd position from the 5' end of the spacer, which was thus revealed to be too close to the 5' end to confer a selective effect with WT Cas9 nuclease(Wang et al., 2019). The Cas9n-subunit of the PE2 fusion protein is derived from the WT nuclease, with a single mutation (H840A) to disable one nuclease domain (Anzalone et al., 2019), and is thus expected to have similar sensitivity to spacer-target mismatches. Several engineered versions of Cas9 have been designed to confer higher target specificity, and are thus more sensitive to mismatches at the 5' end (e.g. eCas9 and Cas9-HF)(Wang et al., 2019). By re-engineering the Cas9n subunit of the PE2 fusion protein to acquire the specificity of these alternatives, sensitivity of the ngRNAs could likely be improved. This theory was not tested as only standard Cas9, and not the enhanced versions (eCas9) was available in the recombinant purified form. However, since plasmids for eCas9 were available, one could in theory co-express eCas9 and the ngRNA plasmid alongside the PE plasmids for *in cellulo* purposes, although it would add another plasmid to an already complex transfection mix.

5.1.2. Prime Editing of rs1799993 in ASC52telo cells

Transfection with chemical reagents

A pilot prime editing experiment targeting rs1799993 was previously attempted in human preadipocytes in our lab, but out of only 2 viable clones, none were successfully edited (Samuelsen, 2021), which is not surprising given that editing efficiencies has been reported to range from 5-60% depending on cell type and their ease of transfection (Anzalone et al., 2019; Grunewald et al., 2023). In this thesis, the AD-MSC cell line ASC52telo was used due to their biological relevance. Unfortunately, these cells proved to be resistant to all tested methods of transfection in combination with the PE2-encoding plasmid, pCMV-PE2-GFP. For chemical transfection methods, the lack of PE2-GFP expression was coupled with an extremely high level of cytotoxicity. This result was confirmed to be cell type-specific, as the plasmid was well-expressed in a control cell line, HEK293T (Figures 4.7 - 4.9).

For successful transfection, the plasmid has to traverse the cell membrane, navigate its way through the cytoplasm (either by diffusion or active transport along microtubules), and subsequently enter the cell nucleus. Cells have natural defense mechanisms against exogenous DNA entry, and for a transfection method to be successful, it has to aid the transport of the DNA molecule into the cell nucleus without triggering these defenses. Chemical reagents, which are generally cationic lipids or polymers, aid in transporting the DNA

across the plasma membrane by forming a complex with the negatively charged DNA(Bruininks et al., 2020). The positively charged headgroups facing the solution interact with the plasma membrane and facilitate cell uptake, usually by endocytosis(Bruininks et al., 2020). Once inside the cell, the complexes may either be actively or passively transported, depending on the nature of the reagent formulation(Cardarelli et al., 2016). Active transport involves the DNA-reagent complexes being captured in endosomes that are transported along microtubules(Cardarelli et al., 2016). The plasmid DNA will then have to escape the endosome before it fuses with the lysosome where foreign DNA is destined for acidic degradation(Bruininks et al., 2020). With passive transport (which is a suggested mechanism for Lipofectamine 2000), the DNA-reagent complex diffuses across the cytosol by Brownian motion, and thus the plasmid DNA has a much higher likelihood of release into the cytoplasm and successfully reaching the cell nucleus (Cardarelli et al., 2016). However, cells have sensing mechanisms to detect DNA presence in the cytosol (Semenova et al., 2019), which can trigger DNA damage repair mechanisms (Piras et al., 2017). Importantly, stem cells are extremely sensitive to DNA damage, which is known to trigger an apoptotic response in these cells (Liu et al., 2014). This makes sense considering that DNA damage in stem cells can potentially be much more detrimental than in differentiated cells. When the apoptotic response is triggered, the cell will self-destruct in a particular manner, in which the cell body shrinks and fragments into apoptotic vesicles (Saraste & Pulkki, 2000). A mode of destruction matching this description was detected in many of the images of transfected ASC52telo cells, in those successfully expressing GFP as well as those not (Figure 5.1). Thus, the observed sensitivity to transfection could potentially be explained by the ASC52telo cells having a heightened susceptibility to apoptosis that is easily triggered by the presence of plasmid DNA in the cytosol. This also agrees with the individual differences in cytotoxicity seen with the various reagents, where those formulated for higher efficiency also produced higher toxicity.

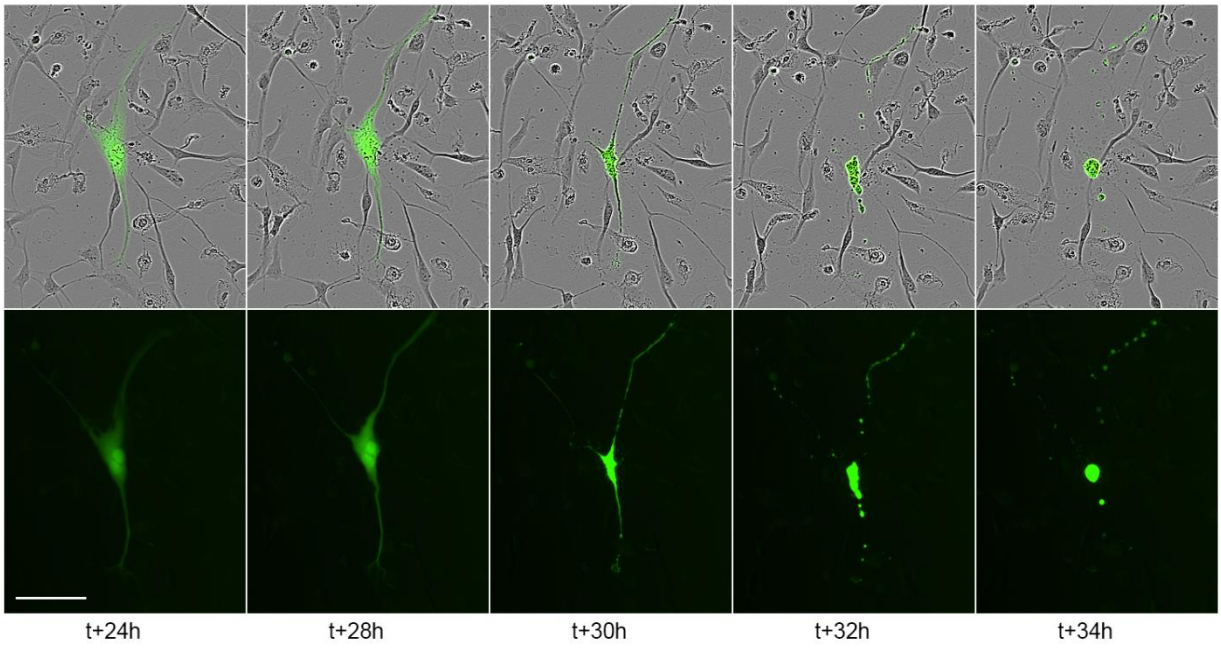


Figure 5.1. | The extremely low viability of chemically transfected ASC52telo cells may be due to transfection triggering an apoptotic response. Here, a GFP-expressing ASC52telo cell is followed over the time span of 10 h. The cell appears to shrink and fragment into contained vesicles; a mode of self-destruction that aligns with that described for apoptosis. Cells were transfected with 60 ng pCAG-eCas9-T2A-GFP plasmid using Lipofectamine Stem Reagent at a DNA-to-reagent ratio of 1:2 (mildest conditions tested). Top panel: phase contrast and green fluorescent channel merged. Bottom panel: green fluorescent channel only. Scale bar = 100 μm.

Further in support of apoptosis being a hindering mechanism for successful transfection lies in the easy transfectability of HEK293T cells. HEK293 cells (and thus their derivative HEK293T) were immortalized by the addition of two genes from adenovirus; E1A and E1B (Berk, 2005). E1A inhibits cell cycle control pathways, but is not sufficient to result in immortality, as it also tends to induce apoptosis. This effect was thus inhibited by the addition of E1B, which encodes two proteins which both inhibit apoptosis through complementary mechanisms. As a by-product of their immortalization, therefore, HEK293 cells also became highly transfectable (Berk, 2005). In line with this thinking, several chemical methods have been suggested that "prime" MSCs for more successful transfection, including adding compounds that inhibit apoptosis to the cell culture medium before and/or after transfection (Hamann et al., 2019). This tactic was not attempted in the current thesis, due to limited knowledge about the use of such compounds and their effect on the stem cell properties of MSCs, as well as potential effects on gene expression. Considering that the aim of transfection is to perform a genetic edit for the detection of downstream changes in gene expression and consequential effect on phenotype including differentiation capacity, it is vital to achieve this without introducing additional variables that may serve as confounders. Therefore, instead of further optimizing transfection with chemical reagents, we sought other methods of gene delivery that might have less cytotoxic effects on the cell

Nucleofection

Where chemical transfection methods have failed, nucleoporation has often proved successful in many cell types, including MSCs (Hamann et al., 2019). We therefore managed to get access to a nucleoporation system to test whether it would improve plasmid delivery to the ASC52 cells. The preliminary test of nucleofection of pmaxGFP in ASC52telo-cells was very promising, as the cells displayed high transfection efficiency, and did not exhibit any of the apoptotic or cytotoxic morphologies seen with chemical reagents. The low cytotoxicity may be a result of the nucleofection method transporting the plasmid directly into the cell nucleus, thus bypassing any DNA-sensing systems in the cytosol (Nakashima et al., 2005). It was therefore somewhat surprising that not a single cell expressed GFP following nucleofection with the pCMV-PE2-GFP plasmid (Figure 4.16). Furthermore, the PE-transfected cells seemed to decrease in confluency during the 48 h following transfection (Figure 4.17), which may indicate that the cells were experiencing additional cytotoxicity apart from that caused by the pulsing (pulse-only controls did not decrease in confluency during the same time-span). These results are likely explained by the massive size of the PE plasmid, as plasmid sizes above 10 kb has been reported to dramatically reduce electroporation efficiencies and cell viability in MSCs (Lesueur et al., 2016). Two pulse-programs were initially tested with the smaller pmaxGFP plasmid, one optimized for higher viability, and the other for higher transfection efficiency, with the former being selected for use with the PE2-plasmid since it yielded the best results. However, it is possible that stronger pulsing (i.e. the program optimized for efficiency at the cost of viability) would have yielded better results with the larger plasmid, and is worth testing in the future. Furthermore, also here "priming" of MSCs prior to transfection by the addition of components to the medium could enhance plasmid delivery to the nucleus. For instance, one study reports that the addition of glucocorticoids (i.e. 100 nM dexamethasone) to the cell medium 30 min prior to transfection has several transfection-enhancing effects, including the dilation of nuclear pores (Kelly et al., 2016), which might enable more efficient entry of larger plasmids. Moreover, priming with glucocorticoids has a reported anti-inflammatory effect that could reduce sensitivity to apoptosis, and has no detected inhibitory effect on differentiation capacity. In fact, the latter was reported to be enhanced (Kelly et al., 2016), which may indicate risk of spontaneous differentiation and should therefore be researched more comprehensively before use.

Considering that the nucleofection method produced high transfection efficiency and viability of surviving cells when transfected with the plasmid of smaller size, minimizing plasmid size could enable PE to be performed in MSCs in the future. This can be achieved by a combination of reducing the length of the vector backbone, along with reducing the length of the transgene. To reduce the vector backbone, the PE2-GFP sequence could be incorporated into a mini-vector, which would substantially decrease the size of the DNA molecule to be delivered to the cell (Hardee et al., 2017). Mini-vectors are designed in such a way that a larger parent plasmid harbors the sequences required for amplification in bacteria (e.g. ori, antibiotic resistance gene, etc) together with the target sequence/expression cassette. The parent plasmid is designed with

recombination sequences that allow it to be converted into two concatenated DNA circles following bacterial amplification and purification (by the addition of lambda integrase); the mini-plasmid containing the bacterial sequences, and the mini-vector with the sequences for transfection. The two catenanes are then unlinked by topoisomerase IV, and the mini-vector isolated by an optimized purification system (Hardee et al., 2017; Hou et al., 2015). However, this would at most decrease the plasmid by ~2000 bp, from 10.5 kb to 8.5 kb, which may not be sufficient for successful transfection. Therefore, a second suggestion is to split the PE2 fusion protein into its respective composite components, Cas9n and RT, such that they can be encoded on two separate plasmids. To facilitate their "fusion" following translation, inspiration could for instance be taken from the CA/I system (Li et al., 2020); by adding two MS2-hairpins to the pegRNA and an MCP-domain to the RT protein, the RT-subunit would be tethered to the Cas9n via the pegRNA. The combination of these two strategies could potentially reduce the size of the largest plasmid required for PE (that encoding dCas9-GFP) to < 6000 bp. In fact, the addition of a linking system may even prove not to be necessary, as recent work has shown that the Cas9 nickase and the RT domain function equally well for PE when they are not physically linked, and can therefore be encoded on two separate, smaller vectors without adversely affecting editing efficiencies (Grunewald et al., 2023).

5.1.3. CRISPR/Cas9-mediated enhancer modulation of 11q23.3 in ASC52telo

Lentiviral transduction

While prime editing can be used to test for both SNP causality and target gene(s), CRISPR/Cas9-mediated enhancer activation/inhibition is an alternative method that can also be used to assess the latter (Li et al., 2020). Since prime editing proved difficult, we turned to this alternative method in order to find the target gene(s) of the 11q23.3 haploblock. While unable to establish SNP causality, this method offers several advantages; while genome editing is permanent, the CA/I system can be designed to be inducible (Li et al., 2020). This is a major advantage as the temporal effect of the enhancer activation can be investigated during mesenchymal stem cell differentiation to mature adipocytes. Moreover, Li et al (2020) designed the system for lentiviral transduction, which has often been found to be superior to transfection methods in difficult-to-transfect-cells, including MSCs (McGinley et al., 2011; Zhou et al., 2023). Thus, the system was adapted for use in ASC52telo cells for the stable integration of the Tet-on-3G and (effector)-dCas9 sequences into the host genome for later induction of the CA/I system.

Despite lentiviral transduction reportedly being a preferable method of gene delivery for MSCs (Zhou et al., 2023), the first round of lentiviral transduction in the current work failed to produce any doubly transduced cells, and scarcely any singly transduced cells (Tables 4.4 and 4.5, respectively). This was also despite the fact that the lentiviral titer tests indicated that sufficiently high levels of the capsid protein p24 were present in the purified and concentrated lentiviral preparation. This, therefore, may indicate that the actual IFU of the harvested virus was much lower than estimated, and the majority of the p24 in the medium belonged

primarily to defective viral particles. p24 is the main protein component of the viral capsid that surrounds the RNA genome (the transfer sequence), with a known number of molecules per virion, and is therefore commonly used to calculate concentration of physical viral particles (Kalidasan et al., 2021). However, the p24 measurement does not indicate number of infectious units (IFU). Successful lentiviral packaging is dependent on multiple factors, and incorrectly assembled particles may lack the transfer sequence, or be otherwise incorrectly assembled and thus not capable of cell entry or integrating the transfer sequence into the host genome (Perry & Rayat, 2021). The effector-dCas9-mCherry transfer sequences are unusually large, at ~10 kb (plasmid size 16 kb), and size of the RNA genome is a known factor that affects packaging efficiency (Kumar et al., 2001; Pirona et al., 2020), with significant decreases seen for every additional kb above ~5 kb (Cante-Barrett et al., 2016). Thus, despite these exact transfer plasmids being sourced from the Li et al. (2020) paper, where they were successfully used for lentiviral transduction (though in other cell types, the implications of which are discussed below), the exact details of the lentiviral packaging or transduction procedures were not provided, and it is possible that the larger genome size produces less efficient incorporation into the virion and should have been accounted for in the estimated titer. Furthermore, it is worth noting that we used other packaging plasmids, and that optimization between packaging plasmids and transfer plasmid my be required to achieve higher percentage of IFUs per PP, as well as optimizing the molar ratios of packaging plasmid to transfer plasmid (Sweeney & Vink, 2021).

Other optimizations to the lentiviral transduction protocol used may also be necessary. For instance, the packaging cell line HEK293T displayed signs of high cytotoxicity in response to the transfection and during packaging; the cells stopped proliferating, became rounded and partially detached from the culture dish, and formed syncytia. Although syncytia formation is a known effect of the VSV-G envelope protein expression in acidic environments (Ci et al., 2018), it is not an expected appearance in successful lentiviral packaging reactions, and its appearance, together with other signs of toxicity, are correlated with low viral titer (Kafri et al., 1999; Ozog et al., 2019; Perry & Rayat, 2021). The lentiviral vectors in the present study were produced in two separate lentiviral packaging reactions (one for each of the planned transduction reactions, respectively), and thus only the first vector production was used. In an attempt to mitigate the observed cytotoxicity of the first round, during the second round the HEK293T cells were exposed to transfection complexes for a reduced length of time (7 h vs 18 h), which did have a visible effect; despite rounding, the cells started to produce syncytia at a later time point, and did not detach from the culture dish. Further improvements may therefore be achieved by using a less toxic transfection reagent. For instance, one study showed that Lipofectamine 3000 produced less toxicity in HEK293T during packaging compared to Lipofectamine 2000 (as was used in the present thesis) and thus higher titer (Kalidasan et al., 2021), while a second study reports that a combination of Lipofectamine 300 and Lipofectamine LTX produces superior lentiviral production in HEK239T than either of the two reagents individually (Pirona et al., 2020). Another low-toxicity option would be to use the TransIT-LT1 reagent, as demonstrated in Figure 4.7.

High toxicity in the packaging cell line can also adversely affect the transduction of the target cells; the toxicity-induced cell death of the packaging cell line may produce cell fragments that are not efficiently removed from the harvested viral preparation, even following filter-purification and up-concentration (Ozog et al., 2019). When exposed to the target cells, cell fragments (and other impurities, including residual DNA and transfection reagents) from the packaging-line may induce further toxicity and hinder successful transduction (Ozog et al., 2019; Perry & Rayat, 2021). It is likely that this occurred in the present study, and, considering the previously discussed sensitivity of ASC52telo cells, may have contributed to the poor transduction outcomes.

Titer of lentiviral vector preparations is known to be affected by freeze-thaw cycles, with up to 30 % loss per cycle (Jiang et al., 2015). The preparations used for the first transduction event in the current work were frozen immediately after harvesting at -80 ºC, and thawed on ice directly prior to use. Since titer was not assessed prior to freezing, quantitative measures of loss are unknown; however, the ideal would be to avoid freezing and use the viral preparations directly. In cases where this is not feasible, steps can be taken to preserve titer during freezing, such as snap-freezing in liquid nitrogen and the addition of stabilizing compounds to the freezing medium (Kumru et al., 2018).

Aspects of the transduction procedure in the target cells may also affect transduction efficiency (see Zhang et al. (2004) for optimization suggestions). Furthermore, different cell lines vary in their inherent susceptibility to transduction (Zhang et al., 2004), which is affected by inherent factors such as level of receptor expression (LDL-receptor for VSV-G pseudotyped lentivirus) (Amirache et al., 2014), presence of cytosolic sensors that recognize and inhibit the viral components (Piras et al., 2017), and factors that prevent or delay capsid uncoating and reverse transcription of the dsRNA into dsDNA for host genome integration (Cosnefroy et al., 2016). This final point is affected by a combination of target cell type and RNA genome size, with the effect being more pronounced in stem cells (Cante-Barrett et al., 2016). Reportedly, even small increases in size above 6 kb can have a significant impact on transduction efficiency, and may therefore be of specific relevance for the current study (Cante-Barrett et al., 2016).

A final note on the lentiviral transduction procedure, is that in current protocols, polybrene is recommended as a transduction enhancer to facilitate virion uptake by the target cells, and this was also used in the present thesis. However, polybrene has been shown to not be the ideal reagent for this purpose in MSCs, as it inhibits proliferations as well as differentiation capacity (Lin et al., 2011; Lin et al., 2012). Indeed, reduced proliferation and change in morphology (thinner and more spindly, data not shown) were observed in the ASC52telo cells following transduction. As discussed previously, it is important to consider the potential downstream effects of the procedures used to limit confounders that may obscure or bias the signal we are looking for. Other agents have been suggested and tested in MSCs (e.g. protamine sulfate) that better preserved these characteristics, and even produced favorable transduction results in MSCs (Collon et al., 2022; Lin et al., 2011; Lin et al., 2012).

Taken together, it is likely that multiple factors contributed to the poor transduction outcome of ASC52telo cells in the current thesis, and that by optimizing the packaging reaction in HEK293T, higher viral titer can be achieved. Furthermore, it would be beneficial to avoid freezing of the viral vectors between harvest and transduction, and optimization of the transduction procedure in ASC52telo may enhance virion uptake and health of the transduced cells. If these measures are not sufficient to achieve the desired outcomes, a reduction in the lengths of the transfer sequences may be required here also, as for PE.

5.2. General discussion

In the 25 years since the first human reference genome was made available to the scientific community, the search for genetic explanations of individual differences in disease susceptibility has greatly intensified, with GWAS as the most applied methodology for detecting genetic loci associated with disease. However, despite the hundreds of thousands of GWAS signals detected for polygenic disease, very few have been successfully linked to mechanism (Alsheikh et al., 2022; Lappalainen & MacArthur, 2021). In part, this discrepancy is due to the lack of well-developed methods for studying non-coding genetic polymorphisms, a category that encompasses over 90 % of the signals associated with polygenic diseases (Lappalainen & MacArthur, 2021). However, with recent breakthroughs in systematic variant-to-function studies, as well as newly developed CRISPR-based tools suitable for *in cellulo* evaluation of potential causal SNPs and affected regulatory elements, we might at last be at a turning point that in the near future, could lead to a dramatic increase in the number of studies successfully linking these signals to causal genetic variations and disease mechanisms.

The fact that humans contain roughly the same number of genes as a simple nematode (C elegans Sequencing Consortium, 1998; Venter et al., 2001), underlines the importance of gene regulation in organismal function. Thus, studying genetic variation in the non-coding DNA is not only an opportunity to discover hitherto unknown disease etiologies, but will likely provide new and fundamental understanding about gene regulatory hierarchy and long-range interactions (Lappalainen & MacArthur, 2021). Within this context, the present study has been conducted as part of a larger project with the aim to systematically dissect a haploblock associated with visceral obesity, as a proof-of-concept variant-to-function study. With previous work having been done to narrow down potential causal SNPs, and suggested affected cell type and potential upstream effectors and downstream target genes, the current aim was to establish the CRISPR/Cas9-derived tools Prime Editing and CA/I, for in situ evaluation of the suggested SNP and surrounding enhancer element in MSCs.

The gold standard of *in cellulo* validation of causal SNPs in a gene regulatory region, is *in situ* genome editing for the downstream assessment of affected gene expression (Claussnitzer et al., 2020). Up until recently, suitable methods available of this type of edit were limited mainly to the CRISPR-mediated homology-directed repair (HDR) gene editing strategy (see Section 1.5), but as previously mentioned, that

method has low efficiency, with a high risk of off-target effects and artefact formation (Anzalone et al., 2019). However, the PE system addresses these problems, as it has significantly higher efficiency and fewer off-targets (Anzalone et al., 2020; Anzalone et al., 2019), and was therefore chosen as the genome-editing strategy for the current work. Of 5 candidate causal SNPs within the predicted enhancer region, rs1799993 was the only one that met the required criteria for editing via the PE system, and was therefore selected as the primary candidate for use with this method in the present study. To safeguard against the conceivable event that the effects of a single edit will not be strong enough to mediate a detectable downstream effect, the CA/I system was also prepared for targeting the 11q23.3 enhancer. Furthermore, this method will serve to strengthen any detected effects of the PE editing.

To avoid potential off-target effects, which would introduce unwanted bias or obscure the signal of interest, a strong emphasis has been placed on the design of spacer sequences for the pegRNA and sgRNAs of the two systems, respectively, which were thus evaluated via multiple predictive algorithms. Furthermore, the candidate sgRNAs were tested *in vitro* on the target template to ensure high on-target efficiency. Similar emphasis was put on the design of the extensions for the pegRNAs. While the editing efficiencies of pegRNAs are not readily assessed *in vitro*, as their function depends on multiple *in vivo* mechanisms (Anzalone et al., 2019), new knowledge and mor in-depth *in silico* evaluation tools are rapidly being developed that can be used for further optimization if necessary in the future (Nelson et al., 2022).

Once the desired edits have been achieved, and the various ASC52telo-modified cell lines established (i.e. homozygous for risk vs protective allele of rs1799993, and/or inducible CA vs CI of the enhancer, respectively) they can be assessed for downstream changes in gene expression. This can be done either in a targeted approach using qPCR against nearby genes, or in an unbiased way using RNA-sequencing. The former is useful as an initial screen to see if the SNP and surrounding enhancer affects the closest genes, such as *HMBS* or *VPS11*. However, as explained in the introduction of this thesis, enhancers may affect genes up to a million bases away (Carullo & Day, 2019), which, in the case of rs1799993, encompasses about 50 genes (Samuelsen, 2021). Therefore RNA-sequencing should also be performed, which detects global changes in gene expression. However, as this would include all consequential changes in gene expression, it is important to keep in mind that the exact nature of the interaction will not be revealed from gene expression data alone. Enhancers may have wide-reaching effects through the regulation of TF-encoding genes for the activation of wider gene-networks (Carullo & Day, 2019), and in such a case it would not be immediately evident which genes are directly affected by the SNP/enhancer, and which are affected as a secondary effect of altered TF levels. Thus, additional strategies will have to be employed to elucidate the sequential interactions and mechanisms. For instance, gene clustering and network analyses may provide important clues on the direction of regulation (Serin et al., 2016). Moreover, since enhancers generally regulate genes on the same chromosome (Carullo & Day, 2019), differential gene expression detected for genes located on other chromosomes than the SNP/enhancer can be attributed to secondary effects. Finally, for robust evidence, chromosome conformation

capture can be utilized to directly demonstrate enhancer-target gene interactions. For this purpose, the 4C sequencing method would be particularly useful, as it can be designed to detect all chromatin loci interacting with a specific DNA element (i.e. the SNP/enhancer) only (Krijger et al., 2020).

After identifying the downstream target genes, the next step would be to infer their function in the cells. By performing gene ontology (GO) and gene pathway enrichment analyses, such as PANTHER analysis (Mi et al., 2013; Thomas et al., 2022), likely affected biological pathways can be detected. This may in turn guide decisions on which functional analyses to perform. For example, if the biological pathway "regulation of fat cell differentiation" would be enriched among differentially expressed genes between homozygous risk and protective carriers at rs1799993, the differentiation capacity of the cells can be tested *in vitro.* If GO analyses don't provide clear clues, standard functional assays related to fat cell biology may be performed, such as assays measuring proliferation and differentiation, mitochondrial function, glucose- and fatty acid uptake, lipolysis etc.

In any genome editing experiment, and particularly for Cas9-related methods, off-target effects are a concern that must be addressed (Zhang et al., 2015). While multiple methods can be used to assess off-targets, deep genome-wide DNA sequencing is a useful and comprehensive way to detect such changes, , although expensive and time-consuming. However, it may be hard to elucidate whether any detected off-target edits have a functional effect on the cells. To address this issue, reversing the edit in the newly edited cell lines could be performed, which should consequently also reverse any observed effects directly related to the SNP, such as gene expression and other observed phenotypes. It is extremely unlikely that off-target effects should be reversed in the same way, so any reversible phenotype can safely be attributed to the primary edit site.

This thesis has a number of strengths; first, genome editing is the gold standard way of validating proposed causal SNPs and is one of the major methods used in this thesis. Secondly, this thesis has put a strong emphasis on guide RNA quality, including specificity and efficiency to reduce potential off-target effects and promote efficient editing. Third, Prime editing has been applied to promote specific editing without double-stranded breaks and reduced off-target effects. Fourth, a complementary method, CRISPR activation/repression has been applied to further strengthen findings from the prime editing. Fifth, these strategies provide an unbiased approach for identifying target genes of interrogated SNP/enhancer. Sixth, the choice of a mesenchymal stem cell model is the most biologically relevant cell model as inferred from epigenetic data.

This thesis also has several limitations: First, the plasmids used for both PE and CA/I are very large, which negatively affects transfection and transduction efficiencies, respectively, currently limiting the usefulness of the methods to easily transfected cell types. Secondly, the selected cell model (ASC52telo), although most biologically relevant, turned out to be technically unsuitable for a proof-of-principle experiment. Thirdly, in terms of answering the biological question related to visceral obesity, an even more precise cell model than ASC52telo could potentially be used, as further explained below.

Despite the fact that the SNP and enhancer region investigated in this study are thought to contribute to the mechanisms underlying *visceral* obesity, the MSC cell line used in this thesis (ASC52telo) are derived from *subcutaneous* adipose tissue. Evidence indicates that MSCs derived from the VAT vs SAT depots have distinct phenotypic characteristics, epigenetic signatures, and gene expression profiles (Chau et al., 2014; Gesta et al., 2006; Macotela et al., 2012). Considering that the phenotypic differences include differentiation capacity (Macotela et al., 2012), which is thought to play a role in the harmful effects of obesity (Kulyte et al., 2022), the source of MSCs used for studying the enhancer need to be taken into account. However, our current understanding of the mechanisms driving VAT vs SAT expansion is limited, and both stem cell populations are thought to potentially contribute. Simplified, increased VAT mass may result from either a reduced differentiation capacity and adipocyte number in SAT, which results in excessive dietary fat being ectopically deposited in VAT (Alligier et al., 2013; Yeoh et al., 2015); or an increased differentiation capacity in VAT causing excessive fat to be more easily deposited in this depot (Van Harmelen et al., 2004). Therefore, the ideal would be to use MSCs derived from both fat depots. However, VAT-derived MSCs are extremely difficult to obtain, due to VAT being less easily biopsied, and because stem cells form VAT are notoriously difficult to culture and differentiate *in vitro* (Macotela et al., 2012), the subcutaneous ASC52telo mesenchymal stem cells is a good candidate for preliminary assessment.

Overall, while the two Cas9-based methods of genetic manipulation tested in this thesis have major advantages as discussed above, the difficulties experienced when applying them to the specific cell type of this thesis highlight the need for newly developed methods to be further optimized to address the individual considerations of various cell types. Once these initial difficulties have been addressed, the components for performing both types of enhancer modulation in ASC52telo cells are prepared and ready for use.

5.3. Conclusions

While the work in the current thesis did not answer the question of whether rs1799993 is a causal SNP in the VAT mass-associated locus 11q23.3, nor identify the target genes of the enhancer within which the SNP is located, the work of this thesis has laid the groundwork for utilizing the PE and CA/I methods to address those questions in the future:

- sgRNA spacers targeting rs1799993 and the surrounding enhancer have been designed with near-ideal predictions of both on-target and off-target effects.
- The sgRNAs have been functionally tested *in vitro* for which the majority exhibited exemplary performance.
- Six different plasmids harboring the various pegRNAs for prime editing of rs1799993 from both risk to protective and vice versa have been cloned, purified and sequence-verified.
- Nine different plasmids harboring the various sgRNAs for CRISPR-mediated activation or repression of the 11q23.3 enhancer have been cloned, purified and sequence-verified.

- The genotype at rs1799993 in ASC52telo cells has been established and found to be heterozygous.
- A comprehensive testing of transfection methods in ASC52telo cells provides invaluable information for further optimization, with nucleofection in combination with smaller plasmid sizes being the most likely successful strategy.
- Lentiviruses for CRISPR-mediated enhancer activation/inhibition were made and a pilot-transduction of ASC52telo cells was performed, suggesting further optimization of viral packaging is required.

Overall, this thesis has contributed towards establishing the genome editing tools required for precise SNP editing and enhancer activation/inhibition and uncovered important areas for further optimizations.

6. Future perspectives

Performing PE and CA/I in MSCs proved to be much more challenging than anticipated, and thus for the successful evaluation of the 11q23.3 locus and its relevance for visceral obesity, as well as elucidating the specific contribution of rs1799993 to disease mechanism, further optimization of the *in cellulo* evaluation methods for use in MSCs will have to be performed.

The hope was that these newly developed CRISPR/Cas9-derived tools would prove to be relatively efficient methods for evaluating the effects of regulatory SNPs on gene expression and cellular phenotypes. However, as evident from the work in this thesis, this turned out not to be the case, as different cell types respond differently to various procedures, and thus newly established methods need to be individually optimized depending on target cell type. In this regard, MSCs proved to be exceptionally challenging, and although many hurdles remain before the PE and CA/I systems can be used successfully in these cells, further optimizing the methods for use in MSCs will likely be well-worth the effort, not only for the current objective, but also for the field as a whole. Dysfunctional gene regulation in stem cells specifically can potentially have far-reaching effects (Carullo & Day, 2019), and thus will likely be an important cell category for the interrogation of SNPs associated with disease. Successful optimization of gene editing tools for use in these sensitive cells will therefore benefit numerous studies in the future.

For PE, the challenge remains of finding a suitable method for the delivery of the PE-encoding components. Nucleofection was shown to produce high transfection efficiency with a smaller plasmid, but not when attempted with the larger PE-plasmid. Thus future optimalization efforts may be directed toward reducing the size of the PE-plasmid, as well as including chemical primers in the cell medium during transfection to facilitate nuclear pore opening as well as reducing the cells' inherent sensitivity to apoptotic triggers (i.e. DNA in cytosol). In the mean time, and considering that all the components required for the edit

of rs179993 have been prepared, the system could potentially be tested in a more easily transfectable cell line for the evaluation of pegRNA editing efficiency, as well as further optimizing the ngRNA design. For CA/I, the problem was most likely a combination of factors relating mainly to the lentiviral packaging, and thus optimization of this procedure will likely produce good results.

Once editing is achieved, the target genes can be assessed by qPCR and/or mRNA sequencing as described above. Further, phenotypic characteristics including effects on metabolism, differentiation capacity, and various other factors thought to play a role in adipose tissue expansion must be performed. A reverse edit will have to be performed to evaluate off-target effects. Finally, an identical evaluation of the SNP and enhancer should also be performed in a VAT-derived MSC cell line, as both VAT and SAT-derived MSCs are thought to potentially influence body fat distribution.

7. References

1000 Genomes Project Consortium, T., Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., Marchini, J. L., McCarthy, S., McVean, G. A., & Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68-74. https://doi.org/10.1038/nature15393

Adams, K. F., Schatzkin, A., Harris, T. B., Kipnis, V., Mouw, T., Ballard-Barbash, R., Hollenbeck, A., & Leitzmann, M. F. (2006). Overweight, obesity, and mortality in a large prospective cohort of persons 50 to 71 years old. *N Engl J Med*, *355*(8), 763-778. https://doi.org/10.1056/NEJMoa055643

Ahmad, S., Rukh, G., Varga, T. V., Ali, A., Kurbasic, A., Shungin, D., Ericson, U., Koivula, R. W., Chu, A. Y., Rose, L. M., Ganna, A., Qi, Q., Stancakova, A., Sandholt, C. H., Elks, C. E., Curhan, G., Jensen, M. K., Tamimi, R. M., Allin, K. H., . . . Franks, P. W. (2013). Gene x physical activity interactions in obesity: combined analysis of 111,421 individuals of European ancestry. *PLoS Genet*, *9*(7), e1003607. https://doi.org/10.1371/journal.pgen.1003607

Alligier, M., Gabert, L., Meugnier, E., Lambert-Porcheron, S., Chanseaume, E., Pilleul, F., Debard, C., Sauvinet, V., Morio, B., Vidal-Puig, A., Vidal, H., & Laville, M. (2013). Visceral fat accumulation during lipid overfeeding is related to subcutaneous adipose tissue characteristics in healthy men. *J Clin Endocrinol Metab*, *98*(2), 802-810. https://doi.org/10.1210/jc.2012-3289

Alsheikh, A. J., Wollenhaupt, S., King, E. A., Reeb, J., Ghosh, S., Stolzenburg, L. R., Tamim, S., Lazar, J., Davis, J. W., & Jacob, H. J. (2022). The landscape of GWAS validation; systematic review identifying 309 validated non-coding variants across 130 human diseases. *BMC Med Genomics*, *15*(1), 74. https://doi.org/10.1186/s12920-022-01216-w

Amirache, F., Levy, C., Costa, C., Mangeot, P. E., Torbett, B. E., Wang, C. X., Negre, D., Cosset, F. L., & Verhoeyen, E. (2014). Mystery solved: VSV-G-LVs do not allow efficient gene transfer into unstimulated T cells, B cells, and HSCs because they lack the LDL receptor. *Blood*, *123*(9), 1422-1424. https://doi.org/10.1182/blood-2013-11-540641

Anderson, E. C., & Novembre, J. (2003). Finding haplotype block boundaries by using the minimum-description-length principle. *Am J Hum Genet*, *73*(2), 336-354. https://doi.org/10.1086/377106

Antonarakis, S. E., & Beckmann, J. S. (2006). Mendelian disorders deserve more attention. *Nat Rev Genet*, *7*(4), 277-282. https://doi.org/10.1038/nrg1826

Anzalone, A. V., Koblan, L. W., & Liu, D. R. (2020). Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. *Nat Biotechnol*, *38*(7), 824-844. https://doi.org/10.1038/s41587-020-0561-9

Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., Chen, P. J., Wilson, C., Newby, G. A., Raguram, A., & Liu, D. R. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*, *576*(7785), 149-157. https://doi.org/10.1038/s41586-019-1711-4

Babu, K., Kathiresan, V., Kumari, P., Newsom, S., Parameshwaran, H. P., Chen, X., Liu, J., Qin, P. Z., & Rajan, R. (2021). Coordinated Actions of Cas9 HNH and RuvC Nuclease Domains Are Regulated by the Bridge Helix and the Target DNA Sequence. *Biochemistry*, *60*(49), 3783-3800. https://doi.org/10.1021/acs.biochem.1c00354

Barroso, I., & McCarthy, M. I. (2019). The Genetic Basis of Metabolic Disease. *Cell*, *177*(1), 146-161. https://doi.org/10.1016/j.cell.2019.02.024

Berk, A. J. (2005). Recent lessons in gene expression, cell cycle control, and cell biology from adenovirus. *Oncogene*, *24*(52), 7673-7685. https://doi.org/10.1038/sj.onc.1209040

Bruininks, B. M., Souza, P. C., Ingolfsson, H., & Marrink, S. J. (2020). A molecular view on the escape of lipoplexed DNA from the endosome. *Elife*, *9*. https://doi.org/10.7554/eLife.52012

C elegans Sequencing Consortium, T. (1998). Genome sequence of the nematode C. elegans: a platform for investigating biology. *Science*, *282*(5396), 2012-2018. https://doi.org/10.1126/science.282.5396.2012

Cante-Barrett, K., Mendes, R. D., Smits, W. K., van Helsdingen-van Wijk, Y. M., Pieters, R., & Meijerink, J. P. (2016). Lentiviral gene transfer into human and murine hematopoietic stem cells: size matters. *BMC Res Notes*, *9*, 312. https://doi.org/10.1186/s13104-016-2118-z

Cardarelli, F., Digiacomo, L., Marchini, C., Amici, A., Salomone, F., Fiume, G., Rossetta, A., Gratton, E., Pozzi, D., & Caracciolo, G. (2016). The intracellular trafficking mechanism of Lipofectamine-based transfection reagents and its implication for gene delivery. *Sci Rep*, *6*, 25879. https://doi.org/10.1038/srep25879

Carullo, N. V. N., & Day, J. J. (2019). Genomic Enhancers in Brain Health and Disease. *Genes (Basel)*, *10*(1). https://doi.org/10.3390/genes10010043

Chau, Y. Y., Bandiera, R., Serrels, A., Martinez-Estrada, O. M., Qing, W., Lee, M., Slight, J., Thornburn, A., Berry, R., McHaffie, S., Stimson, R. H., Walker, B. R., Chapuli, R. M., Schedl, A., & Hastie, N. (2014). Visceral and subcutaneous fat have different origins and evidence supports a mesothelial source. *Nat Cell Biol*, *16*(4), 367-375. https://doi.org/10.1038/ncb2922

Chretien, S., Dubart, A., Beaupain, D., Raich, N., Grandchamp, B., Rosa, J., Goossens, M., & Romeo, P. H. (1988). Alternative transcription and splicing of the human porphobilinogen deaminase gene result either in tissue-specific or in housekeeping expression. *Proc Natl Acad Sci U S A*, *85*(1), 6-10. https://doi.org/10.1073/pnas.85.1.6

Ci, Y., Yang, Y., Xu, C., & Shi, L. (2018). Vesicular stomatitis virus G protein transmembrane region is crucial for the hemi-fusion to full fusion transition. *Sci Rep*, *8*(1), 10669. https://doi.org/10.1038/s41598-018-28868-y

Claringbould, A., & Zaugg, J. B. (2021). Enhancers in disease: molecular basis and emerging treatment strategies. *Trends Mol Med*, *27*(11), 1060-1073. https://doi.org/10.1016/j.molmed.2021.07.012

Claussnitzer, M., Cho, J. H., Collins, R., Cox, N. J., Dermitzakis, E. T., Hurles, M. E., Kathiresan, S., Kenny, E. E., Lindgren, C. M., MacArthur, D. G., North, K. N., Plon, S. E., Rehm, H. L., Risch, N., Rotimi, C. N., Shendure, J., Soranzo, N., & McCarthy, M. I. (2020). A brief history of human disease genetics. *Nature*, *577*(7789), 179-189. https://doi.org/10.1038/s41586-019-1879-7

Claussnitzer, M., Dankel, S. N., Kim, K. H., Quon, G., Meuleman, W., Haugen, C., Glunk, V., Sousa, I. S., Beaudry, J. L., Puviindran, V., Abdennur, N. A., Liu, J., Svensson, P. A., Hsu, Y. H., Drucker, D. J., Mellgren, G., Hui, C. C., Hauner, H., & Kellis, M. (2015). FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. *N Engl J Med*, *373*(10), 895-907. https://doi.org/10.1056/NEJMoa1502214

Claussnitzer, M., Dankel, S. N., Klocke, B., Grallert, H., Glunk, V., Berulava, T., Lee, H., Oskolkov, N., Fadista, J., Ehlers, K., Wahl, S., Hoffmann, C., Qian, K., Ronn, T., Riess, H., Muller-Nurasyid, M., Bretschneider, N., Schroeder, T., Skurk, T., . . . Laumen, H. (2014). Leveraging cross-species transcription factor binding site patterns: from diabetes risk loci to disease mechanisms. *Cell*, *156*(1-2), 343-358. https://doi.org/10.1016/j.cell.2013.10.058

Collaboration, N. C. D. R. F. (2016). Trends in adult body-mass index in 200 countries from 1975 to 2014: a pooled analysis of 1698 population-based measurement studies with 19.2 million participants. *Lancet*, *387*(10026), 1377-1396. https://doi.org/10.1016/S0140-6736(16)30054-X

Collon, K., Gallo, M. C., Bell, J. A., Chang, S. W., Rodman, J. C. S., Sugiyama, O., Kohn, D. B., & Lieberman, J. R. (2022). Improving Lentiviral Transduction of Human Adipose-Derived Mesenchymal Stem Cells. *Hum Gene Ther*, *33*(23-24), 1260-1268. https://doi.org/10.1089/hum.2022.117

Cosnefroy, O., Murray, P. J., & Bishop, K. N. (2016). HIV-1 capsid uncoating initiates after the first strand transfer of reverse transcription. *Retrovirology*, *13*(1), 58. https://doi.org/10.1186/s12977-016-0292-7

Dey, K. K., Gazal, S., van de Geijn, B., Kim, S. S., Nasser, J., Engreitz, J. M., & Price, A. L. (2022). SNP-to-gene linking strategies reveal contributions of enhancer-related and candidate master-regulator genes to autoimmune disease. *Cell Genom*, *2*(7). https://doi.org/10.1016/j.xgen.2022.100145

Di Angelantonio, E., Bhupathiraju, S. N., Hu, F. B., Danesh, J., Peto, R., Lewington, S., & Global, B. M. I. M. C. (2017). Body-mass index and all-cause mortality - Authors' reply. *Lancet*, *389*(10086), 2285-2286. https://doi.org/10.1016/S0140-6736(17)31369-7

Durand, S., & Cimarelli, A. (2011). The inside out of lentiviral vectors. *Viruses*, *3*(2), 132-159. https://doi.org/10.3390/v3020132

Engler, C., Kandzia, R., & Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. *PLoS One*, *3*(11), e3647. https://doi.org/10.1371/journal.pone.0003647

Frayling, T. M., Timpson, N. J., Weedon, M. N., Zeggini, E., Freathy, R. M., Lindgren, C. M., Perry, J. R., Elliott, K. S., Lango, H., Rayner, N. W., Shields, B., Harries, L. W., Barrett, J. C., Ellard, S., Groves, C. J., Knight, B., Patch, A. M., Ness, A. R., Ebrahim, S., . . . McCarthy, M. I. (2007). A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science*, *316*(5826), 889-894. https://doi.org/10.1126/science.1141634

German Advisory Committee Blood, S. A. o. P. T. b. B. (2016). Human Immunodeficiency Virus (HIV). *Transfus Med Hemother*, *43*(3), 203-222. https://doi.org/10.1159/000445852

Gesta, S., Bluher, M., Yamamoto, Y., Norris, A. W., Berndt, J., Kralisch, S., Boucher, J., Lewis, C., & Kahn, C. R. (2006). Evidence for a role of developmental genes in the origin of obesity and body fat distribution. *Proc Natl Acad Sci U S A*, *103*(17), 6676-6681. https://doi.org/10.1073/pnas.0601752103

Glazier, A. M., Nadeau, J. H., & Aitman, T. J. (2002). Finding genes that underlie complex traits. *Science*, *298*(5602), 2345-2349. https://doi.org/10.1126/science.1076641

Grunewald, J., Miller, B. R., Szalay, R. N., Cabeceiras, P. K., Woodilla, C. J., Holtz, E. J. B., Petri, K., & Joung, J. K. (2023). Engineered CRISPR prime editors with compact, untethered reverse transcriptases. *Nat Biotechnol*, *41*(3), 337-343. https://doi.org/10.1038/s41587-022-01473-1

GTEx Consortium, T. (2013). The Genotype-Tissue Expression (GTEx) project. *Nat Genet*, *45*(6), 580-585. https://doi.org/10.1038/ng.2653

Haeussler, M., Schonig, K., Eckert, H., Eschstruth, A., Mianne, J., Renaud, J. B., Schneider-Maunoury, S., Shkumatava, A., Teboul, L., Kent, J., Joly, J. S., & Concordet, J. P. (2016). Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol*, *17*(1), 148. https://doi.org/10.1186/s13059-016-1012-2

Hamann, A., Nguyen, A., & Pannier, A. K. (2019). Nucleic acid delivery to mesenchymal stem cells: a review of nonviral methods and applications. *J Biol Eng*, *13*, 7. https://doi.org/10.1186/s13036-019-0140-0

Hardee, C. L., Arevalo-Soliz, L. M., Hornstein, B. D., & Zechiedrich, L. (2017). Advances in Non-Viral DNA Vectors for Gene Therapy. *Genes (Basel)*, *8*(2). https://doi.org/10.3390/genes8020065

Hardies, K., Weckhuysen, S., De Jonghe, P., & Suls, A. (2016). Lessons learned from gene identification studies in Mendelian epilepsy disorders. *Eur J Hum Genet*, *24*(7), 961-967. https://doi.org/10.1038/ejhg.2015.251

Herman, M. A., & Rosen, E. D. (2015). Making Biological Sense of GWAS Data: Lessons from the FTO Locus. *Cell Metab*, *22*(4), 538-539. https://doi.org/10.1016/j.cmet.2015.09.018

Hood, L., & Rowen, L. (2013). The Human Genome Project: big science transforms biology and medicine. *Genome Med*, *5*(9), 79. https://doi.org/10.1186/gm483

Hou, X. H., Guo, X. Y., Chen, Y., He, C. Y., & Chen, Z. Y. (2015). Increasing the minicircle DNA purity using an enhanced triplex DNA technology to eliminate DNA contaminants. *Mol Ther Methods Clin Dev*, *1*, 14062. https://doi.org/10.1038/mtm.2014.62

Hsu, J. Y., Grunewald, J., Szalay, R., Shih, J., Anzalone, A. V., Lam, K. C., Shen, M. W., Petri, K., Liu, D. R., Joung, J. K., & Pinello, L. (2021). PrimeDesign software for rapid and simplified design of prime editing guide RNAs. *Nat Commun*, *12*(1), 1034. https://doi.org/10.1038/s41467-021-21337-7

Iacobini, C., Pugliese, G., Blasetti Fantauzzi, C., Federici, M., & Menini, S. (2019). Metabolically healthy versus metabolically unhealthy obesity. *Metabolism*, *92*, 51-60. https://doi.org/10.1016/j.metabol.2018.11.009

International HapMap, C. (2003). The International HapMap Project. *Nature*, *426*(6968), 789-796. https://doi.org/10.1038/nature02168

International HapMap, C. (2005). A haplotype map of the human genome. *Nature*, *437*(7063), 1299-1320. https://doi.org/10.1038/nature04226

International Human Genome Sequencing, C. (2004). Finishing the euchromatic sequence of the human genome. *Nature*, *431*(7011), 931-945. https://doi.org/10.1038/nature03001

Jiang, W., Hua, R., Wei, M., Li, C., Qiu, Z., Yang, X., & Zhang, C. (2015). An optimized method for high-titer lentivirus preparations without ultracentrifugation. *Sci Rep*, *5*, 13875. https://doi.org/10.1038/srep13875

Kafri, T., van Praag, H., Ouyang, L., Gage, F. H., & Verma, I. M. (1999). A packaging cell line for lentivirus vectors. *J Virol*, *73*(1), 576-584. https://doi.org/10.1128/JVI.73.1.576-584.1999

Kalidasan, V., Ng, W. H., Ishola, O. A., Ravichantar, N., Tan, J. J., & Das, K. T. (2021). A guide in lentiviral vector production for hard-to-transfect cells, using cardiac-derived c-kit expressing cells as a model system. *Sci Rep*, *11*(1), 19265. https://doi.org/10.1038/s41598-021-98657-7

Karlsson, T., Rask-Andersen, M., Pan, G., Hoglund, J., Wadelius, C., Ek, W. E., & Johansson, A. (2019). Contribution of genetics to visceral adiposity and its relation to cardiovascular and metabolic disease. *Nat Med*, *25*(9), 1390-1395. https://doi.org/10.1038/s41591-019-0563-7

Kelly, A. M., Plautz, S. A., Zempleni, J., & Pannier, A. K. (2016). Glucocorticoid Cell Priming Enhances Transfection Outcomes in Adult Human Mesenchymal Stem Cells. *Mol Ther*, *24*(2), 331-341. https://doi.org/10.1038/mt.2015.195

Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., & Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, *533*(7603), 420-424. https://doi.org/10.1038/nature17946

Konstantakos, V., Nentidis, A., Krithara, A., & Paliouras, G. (2022). CRISPR-Cas9 gRNA efficiency prediction: an overview of predictive tools and the role of deep learning. *Nucleic Acids Res*, *50*(7), 3616-3637. https://doi.org/10.1093/nar/gkac192

Krijger, P. H. L., Geeven, G., Bianchi, V., Hilvering, C. R. E., & de Laat, W. (2020). 4C-seq from beginning to end: A detailed protocol for sample preparation and data analysis. *Methods*, *170*, 17-32. https://doi.org/10.1016/j.ymeth.2019.07.014

Kulyte, A., Aman, A., Strawbridge, R. J., Arner, P., & Dahlman, I. A. (2022). Genome-Wide Association Study Identifies Genetic Loci Associated With Fat Cell Number and Overlap With Genetic Risk Loci for Type 2 Diabetes. *Diabetes*, *71*(6), 1350-1362. https://doi.org/10.2337/db21-0804

Kumar, M., Keller, B., Makalou, N., & Sutton, R. E. (2001). Systematic determination of the packaging limit of lentiviral vectors. *Hum Gene Ther*, *12*(15), 1893-1905. https://doi.org/10.1089/104303401753153947

Kumru, O. S., Wang, Y., Gombotz, C. W. R., Kelley-Clarke, B., Cieplak, W., Kim, T., Joshi, S. B., & Volkin, D. B. (2018). Physical Characterization and Stabilization of a Lentiviral Vector Against Adsorption and Freeze-Thaw. *J Pharm Sci*, *107*(11), 2764-2774. https://doi.org/10.1016/j.xphs.2018.07.010

Kwok, K. H., Lam, K. S., & Xu, A. (2016). Heterogeneity of white adipose tissue: molecular basis and clinical implications. *Exp Mol Med*, *48*(3), e215. https://doi.org/10.1038/emm.2016.5

Lappalainen, T., & MacArthur, D. G. (2021). From variant to function in human disease genetics. *Science*, *373*(6562), 1464-1468. https://doi.org/10.1126/science.abi8207

Lesueur, L. L., Mir, L. M., & Andre, F. M. (2016). Overcoming the Specific Toxicity of Large Plasmids Electrotransfer in Primary Cells In Vitro. *Mol Ther Nucleic Acids*, *5*(3), e291. https://doi.org/10.1038/mtna.2016.4

Li, K., Liu, Y., Cao, H., Zhang, Y., Gu, Z., Liu, X., Yu, A., Kaphle, P., Dickerson, K. E., Ni, M., & Xu, J. (2020). Interrogation of enhancer function by enhancer-targeting CRISPR epigenetic editing. *Nat Commun*, *11*(1), 485. https://doi.org/10.1038/s41467-020-14362-5

Lin, P., Correa, D., Lin, Y., & Caplan, A. I. (2011). Polybrene inhibits human mesenchymal stem cell proliferation during lentiviral transduction. *PLoS One*, *6*(8), e23891. https://doi.org/10.1371/journal.pone.0023891

Lin, P., Lin, Y., Lennon, D. P., Correa, D., Schluchter, M., & Caplan, A. I. (2012). Efficient lentiviral transduction of human mesenchymal stem cells that preserves proliferation and differentiation capabilities. *Stem Cells Transl Med*, *1*(12), 886-897. https://doi.org/10.5966/sctm.2012-0086

Lin, Y. C., Boone, M., Meuris, L., Lemmens, I., Van Roy, N., Soete, A., Reumers, J., Moisse, M., Plaisance, S., Drmanac, R., Chen, J., Speleman, F., Lambrechts, D., Van de Peer, Y., Tavernier, J., & Callewaert, N. (2014). Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nat Commun*, *5*, 4767. https://doi.org/10.1038/ncomms5767

Liu, J. C., Lerou, P. H., & Lahav, G. (2014). Stem cells: balancing resistance and sensitivity to DNA damage. *Trends Cell Biol*, *24*(5), 268-274. https://doi.org/10.1016/j.tcb.2014.03.002

Locke, A. E., Kahali, B., Berndt, S. I., Justice, A. E., Pers, T. H., Day, F. R., Powell, C., Vedantam, S., Buchkovich, M. L., Yang, J., Croteau-Chonka, D. C., Esko, T., Fall, T., Ferreira, T., Gustafsson, S., Kutalik, Z., Luan, J., Magi, R., Randall, J. C., . . . Speliotes, E. K. (2015). Genetic studies of body mass index yield new insights for obesity biology. *Nature*, *518*(7538), 197-206. https://doi.org/10.1038/nature14177

Macotela, Y., Emanuelli, B., Mori, M. A., Gesta, S., Schulz, T. J., Tseng, Y. H., & Kahn, C. R. (2012). Intrinsic differences in adipocyte precursor cells from different white fat depots. *Diabetes*, *61*(7), 1691-1699. https://doi.org/10.2337/db11-1753

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., & Church, G. M. (2013). RNA-guided human genome engineering via Cas9. *Science*, *339*(6121), 823-826. https://doi.org/10.1126/science.1232033

Manuck, S. B., & McCaffery, J. M. (2014). Gene-environment interaction. *Annu Rev Psychol*, *65*, 41-70. https://doi.org/10.1146/annurev-psych-010213-115100

Maurano, M. T., Humbert, R., Rynes, E., Thurman, R. E., Haugen, E., Wang, H., Reynolds, A. P., Sandstrom, R., Qu, H., Brody, J., Shafer, A., Neri, F., Lee, K., Kutyavin, T., Stehling-Sun, S., Johnson, A. K., Canfield, T. K., Giste, E., Diegel, M., . . . Stamatoyannopoulos, J. A. (2012). Systematic localization of common disease-associated variation in regulatory DNA. *Science*, *337*(6099), 1190-1195. https://doi.org/10.1126/science.1222794

McGinley, L., McMahon, J., Strappe, P., Barry, F., Murphy, M., O'Toole, D., & O'Brien, T. (2011). Lentiviral vector mediated modification of mesenchymal stem cells & enhanced survival in an in vitro model of ischaemia. *Stem Cell Res Ther*, *2*(2), 12. https://doi.org/10.1186/scrt53

Mi, H., Muruganujan, A., & Thomas, P. D. (2013). PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res*, *41*(Database issue), D377-386. https://doi.org/10.1093/nar/gks1118

Mirza, S. S. (2022). *Genotype dependent gene regulation in visceral adiposity: a study of the 11q23.3 locus* [Master's Thesis, University of Bergen].

Moreno-Navarrete, J. M., Rodriguez, A., Ortega, F., Becerril, S., Girones, J., Sabater-Masdeu, M., Latorre, J., Ricart, W., Fruhbeck, G., & Fernandez-Real, J. M. (2017). Heme Biosynthetic Pathway is Functionally Linked to Adipogenesis via Mitochondrial Respiratory Activity. *Obesity (Silver Spring)*, *25*(10), 1723-1733. https://doi.org/10.1002/oby.21956

Nakashima, S., Matsuyama, Y., Nitta, A., Sakai, Y., & Ishiguro, N. (2005). Highly efficient transfection of human marrow stromal cells by nucleofection. *Transplant Proc*, *37*(5), 2290-2292. https://doi.org/10.1016/j.transproceed.2005.03.047

Nelson, J. W., Randolph, P. B., Shen, S. P., Everette, K. A., Chen, P. J., Anzalone, A. V., An, M., Newby, G. A., Chen, J. C., Hsu, A., & Liu, D. R. (2022). Engineered pegRNAs improve prime editing efficiency. *Nat Biotechnol*, *40*(3), 402-410. https://doi.org/10.1038/s41587-021-01039-7

Ng, M., Fleming, T., Robinson, M., Thomson, B., Graetz, N., Margono, C., Mullany, E. C., Biryukov, S., Abbafati, C., Abera, S. F., Abraham, J. P., Abu-Rmeileh, N. M., Achoki, T., AlBuhairan, F. S., Alemu, Z. A., Alfonso, R., Ali, M. K., Ali, R., Guzman, N. A., . . . Gakidou, E. (2014). Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet*, *384*(9945), 766-781. https://doi.org/10.1016/S0140-6736(14)60460-8

Ozog, S., Chen, C. X., Simpson, E., Garijo, O., Timberlake, N. D., Minder, P., Verhoeyen, E., & Torbett, B. E. (2019). CD46 Null Packaging Cell Line Improves Measles Lentiviral Vector Production and Gene Delivery to Hematopoietic Stem and Progenitor Cells. *Mol Ther Methods Clin Dev*, *13*, 27-39. https://doi.org/10.1016/j.omtm.2018.11.006

Perry, C., & Rayat, A. (2021). Lentiviral Vector Bioprocessing. *Viruses*, *13*(2). https://doi.org/10.3390/v13020268

Piras, F., Riba, M., Petrillo, C., Lazarevic, D., Cuccovillo, I., Bartolaccini, S., Stupka, E., Gentner, B., Cittaro, D., Naldini, L., & Kajaste-Rudnitski, A. (2017). Lentiviral vectors escape innate sensing but trigger p53 in human hematopoietic stem and progenitor cells. *EMBO Mol Med*, *9*(9), 1198-1211. https://doi.org/10.15252/emmm.201707922

Pirona, A. C., Oktriani, R., Boettcher, M., & Hoheisel, J. D. (2020). Process for an efficient lentiviral cell transduction. *Biol Methods Protoc*, *5*(1), bpaa005. https://doi.org/10.1093/biomethods/bpaa005

Ranganathan, V., Wahlin, K., Maruotti, J., & Zack, D. J. (2014). Expansion of the CRISPR-Cas9 genome targeting space through the use of H1 promoter-expressed guide RNAs. *Nat Commun*, *5*, 4516. https://doi.org/10.1038/ncomms5516

Riesenberg, S., Helmbrecht, N., Kanis, P., Maricic, T., & Paabo, S. (2022). Improved gRNA secondary structures allow editing of target sites resistant to CRISPR-Cas9 cleavage. *Nat Commun*, *13*(1), 489. https://doi.org/10.1038/s41467-022-28137-7

Roadmap Epigenomics, C., Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M. J., Amin, V., Whitaker, J. W., Schultz, M. D., Ward, L. D., Sarkar, A., Quon, G., Sandstrom, R. S., Eaton, M. L., . . . Kellis, M. (2015). Integrative analysis of 111 reference human epigenomes. *Nature*, *518*(7539), 317-330. https://doi.org/10.1038/nature14248

Samuelsen, N. T. (2021). *Functional dissection of a genetic locus for visceral fat mass (11q23.3)* [Master's Thesis, University of Bergen].

Saraste, A., & Pulkki, K. (2000). Morphologic and biochemical hallmarks of apoptosis. *Cardiovasc Res*, *45*(3), 528-537. https://doi.org/10.1016/s0008-6363(99)00384-3

Schwarz, E. J., Reginato, M. J., Shao, D., Krakow, S. L., & Lazar, M. A. (1997). Retinoic acid blocks adipogenesis by inhibiting C/EBPbeta-mediated transcription. *Mol Cell Biol*, *17*(3), 1552-1561. https://doi.org/10.1128/MCB.17.3.1552

Scott, L. J., Mohlke, K. L., Bonnycastle, L. L., Willer, C. J., Li, Y., Duren, W. L., Erdos, M. R., Stringham, H. M., Chines, P. S., Jackson, A. U., Prokunina-Olsson, L., Ding, C. J., Swift, A. J., Narisu, N., Hu, T., Pruim, R., Xiao, R., Li, X. Y., Conneely, K. N., . . . Boehnke, M. (2007). A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science*, *316*(5829), 1341-1345. https://doi.org/10.1126/science.1142382

Semenova, N., Bosnjak, M., Markelc, B., Znidar, K., Cemazar, M., & Heller, L. (2019). Multiple cytosolic DNA sensors bind plasmid DNA after transfection. *Nucleic Acids Res*, *47*(19), 10235-10246. https://doi.org/10.1093/nar/gkz768

Serin, E. A., Nijveen, H., Hilhorst, H. W., & Ligterink, W. (2016). Learning from Co-expression Networks: Possibilities and Challenges. *Front Plant Sci*, *7*, 444. https://doi.org/10.3389/fpls.2016.00444

Sigma Aldrich, M. (n.d). *Successful Transduction Using Lentivirus*. Retrieved 02.06.2023 from https://www.sigmaaldrich.com/NO/en/technical-documents/technical-article/genomics/advanced-gene-editing/successful-transduction-lentivirus

Sinnott-Armstrong, N., Sousa, I. S., Laber, S., Rendina-Ruedy, E., Nitter Dankel, S. E., Ferreira, T., Mellgren, G., Karasik, D., Rivas, M., Pritchard, J., Guntur, A. R., Cox, R. D., Lindgren, C. M., Hauner, H., Sallari, R., Rosen, C. J., Hsu, Y. H., Lander, E. S., Kiel, D. P., & Claussnitzer, M. (2021). A regulatory variant at 3q21.1 confers an increased pleiotropic risk for hyperglycemia and altered bone mineral density. *Cell Metab*, *33*(3), 615-628 e613. https://doi.org/10.1016/j.cmet.2021.01.001

Sladek, R., Rocheleau, G., Rung, J., Dina, C., Shen, L., Serre, D., Boutin, P., Vincent, D., Belisle, A., Hadjadj, S., Balkau, B., Heude, B., Charpentier, G., Hudson, T. J., Montpetit, A., Pshezhetsky, A. V., Prentki, M., Posner, B. I., Balding, D. J., . . . Froguel, P. (2007). A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature*, *445*(7130), 881-885. https://doi.org/10.1038/nature05616

Slatkin, M. (2008). Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. *Nat Rev Genet*, *9*(6), 477-485. https://doi.org/10.1038/nrg2361

Spicuglia, S., & Vanhille, L. (2012). Chromatin signatures of active enhancers. *Nucleus*, *3*(2), 126-131. https://doi.org/10.4161/nucl.19232

Sun, C., Kovacs, P., & Guiu-Jurado, E. (2021). Genetics of Body Fat Distribution: Comparative Analyses in Populations with European, Asian and African Ancestries. *Genes (Basel)*, *12*(6). https://doi.org/10.3390/genes12060841

Sweeney, N. P., & Vink, C. A. (2021). The impact of lentiviral vector genome size and producer cell genomic to gag-pol mRNA ratios on packaging efficiency and titre. *Mol Ther Methods Clin Dev*, *21*, 574-584. https://doi.org/10.1016/j.omtm.2021.04.007

Saad, R. K., Ghezzawi, M., Horanieh, R., Khamis, A. M., Saunders, K. H., Batsis, J. A., & Chakhtoura, M. (2022). Abdominal Visceral Adipose Tissue and All-Cause Mortality: A Systematic Review. *Front Endocrinol (Lausanne)*, *13*, 922931. https://doi.org/10.3389/fendo.2022.922931

Thomas, P. D., Ebert, D., Muruganujan, A., Mushayahama, T., Albou, L. P., & Mi, H. (2022). PANTHER: Making genome-scale phylogenetics accessible to all. *Protein Sci*, *31*(1), 8-22. https://doi.org/10.1002/pro.4218

Thomas, P. D., & Kejariwal, A. (2004). Coding single-nucleotide polymorphisms associated with complex vs. Mendelian disease: evolutionary evidence for differences in molecular effects. *Proc Natl Acad Sci U S A*, *101*(43), 15398-15403. https://doi.org/10.1073/pnas.0404380101

Uffelmann, E., Huang, Q. Q., Munung, N. S., de Vries, J., Okada, Y., Martin, A. R., Martin, H. C., Lappalainen, T., & Posthuma, D. (2021). Genome-wide association studies. *Nature Reviews Methods Primers*, *1*(1). https://doi.org/10.1038/s43586-021-00056-9

Uffelmann, E., & Posthuma, D. (2021). Emerging Methods and Resources for Biological Interrogation of Neuropsychiatric Polygenic Signal. *Biol Psychiatry*, *89*(1), 41-53. https://doi.org/10.1016/j.biopsych.2020.05.022

Van Harmelen, V., Rohrig, K., & Hauner, H. (2004). Comparison of proliferation and differentiation capacity of human adipocyte precursor cells from the omental and subcutaneous adipose tissue depot of obese subjects. *Metabolism*, *53*(5), 632-637. https://doi.org/10.1016/j.metabol.2003.11.012

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., . . . Zhu, X. (2001). The sequence of the human genome. *Science*, *291*(5507), 1304-1351. https://doi.org/10.1126/science.1058040

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet*, *101*(1), 5-22. https://doi.org/10.1016/j.ajhg.2017.06.005

Wallace, C., Dobson, R. J., Munroe, P. B., & Caulfield, M. J. (2007). Information capture using SNPs from HapMap and whole-genome chips differs in a sample of inflammatory and cardiovascular gene-centric regions from genome-wide estimates. *Genome Res*, *17*(11), 1596-1602. https://doi.org/10.1101/gr.5996407

Walsh, P. S., Metzger, D. A., & Higushi, R. (2013). Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. BioTechniques 10(4): 506-13 (April 1991). *Biotechniques*, *54*(3), 134-139. https://doi.org/10.2144/000114018

Wang, D., Zhang, C., Wang, B., Li, B., Wang, Q., Liu, D., Wang, H., Zhou, Y., Shi, L., Lan, F., & Wang, Y. (2019). Optimized CRISPR guide RNA design for two high-fidelity Cas9 variants by deep learning. *Nat Commun*, *10*(1), 4284. https://doi.org/10.1038/s41467-019-12281-8

WHO, W. H. O. (2015). Obesity and Overweight Key Facts. In *Fact Sheet No. 311*.

Wolbank, S., Stadler, G., Peterbauer, A., Gillich, A., Karbiener, M., Streubel, B., Wieser, M., Katinger, H., van Griensven, M., Redl, H., Gabriel, C., Grillari, J., & Grillari-Voglauer, R. (2009). Telomerase immortalized human amnion- and adipose-derived mesenchymal stem cells: maintenance of differentiation and immunomodulatory characteristics. *Tissue Eng Part A*, *15*(7), 1843-1854. https://doi.org/10.1089/ten.tea.2008.0205

Yashin, A. I., Wu, D., Arbeev, K. G., & Ukraintseva, S. V. (2012). Polygenic effects of common single-nucleotide polymorphisms on life span: when association meets causality. *Rejuvenation Res*, *15*(4), 381-394. https://doi.org/10.1089/rej.2011.1257

Yeoh, A. J., Pedley, A., Rosenquist, K. J., Hoffmann, U., & Fox, C. S. (2015). The Association Between Subcutaneous Fat Density and the Propensity to Store Fat Viscerally. *J Clin Endocrinol Metab*, *100*(8), E1056-1064. https://doi.org/10.1210/jc.2014-4032

Zhang, B., Metharom, P., Jullie, H., Ellem, K. A., Cleghorn, G., West, M. J., & Wei, M. Q. (2004). The significance of controlled conditions in lentiviral vector titration and in the use of multiplicity of infection (MOI) for predicting gene transfer events. *Genet Vaccines Ther*, *2*(1), 6. https://doi.org/10.1186/1479-0556-2-6

Zhang, X. H., Tee, L. Y., Wang, X. G., Huang, Q. S., & Yang, S. H. (2015). Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. *Mol Ther Nucleic Acids*, *4*(11), e264. https://doi.org/10.1038/mtna.2015.37

Zhou, H., He, Y., Xiong, W., Jing, S., Duan, X., Huang, Z., Nahal, G. S., Peng, Y., Li, M., Zhu, Y., & Ye, Q. (2023). MSC based gene delivery methods and strategies improve the therapeutic efficacy of neurological diseases. *Bioact Mater*, *23*, 409-437. https://doi.org/10.1016/j.bioactmat.2022.11.007

8. Supplementary information:

Sequences for pegRNA assembly of the pegRNA sequences peg1 (Both A and C versions), peg2 and peg3 (versions for editing from protective to risk allele shown, for reverse edit, the highlighted red nucleotides of the extensions were exchanged to the protective allele (C on top strand, G on bottom). All pegRNAs were assembled from three segments; the spacer sequence, the scaffold sequence, and the extension, by GG-assembly. GG overhangs are indicated for each sequence in lowercase font. For some extensions, RT template is indicated in bold (with edit SNP in lowercase, red) and PBS in cursive. All sequences 5' → 3'.

pegRNA scaffold sequence	5' → 3'
pegRNA scaffold sequence top	AGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCG
pegRNA scaffold sequence bottom	GCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAG

peg1_A	**Edits rs1799993 from protective to risk** Sequence (5' → 3')
Spacer sequence (top)	caccGAAACCAGAGGCAGCACTCTgtttt
Spacer sequence (bottom)	ctctaaaacAGAGTGCTGCCTCTGGTTTC
Extension oligo (top)	gtgc**TTCGTtCACCTAG***AGTGCTGCCTCTGGTT*
Extension oligo (bottom)	aaaa*AACCAGAGGCAGCACT***CTAGGTGaACGAA**
Compiled sequence (top)	caccGAAACCAGAGGCAGCACTCTgttttAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGgtgcTTCGT**t**CACCTAGAGTGCTGCCTCTGGTT
Compiled sequence (bottom)	aaaaAACCAGAGGCAGCACTCTAGGTG**a**ACGAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGctctaaaacAGAGTGCTGCCTCTGGTTTC

peg1_C	**Edits rs1799993 from risk to protective** Sequence (5’ → 3’)
Spacer sequence (top)	caccGAAACCAGAGGCAGCACTCTgtttt
Spacer sequence (bottom)	ctctaaaacAGAGTGCTGCCTCTGGTTTC
Extension oligo (top)	gtgc**TTCGTgCACCTAG***AGTGCTGCCTCTGGTT*
Extension oligo (bottom)	aaaa*AACCAGAGGCAGCACT***CTAGGTGcACGAA**
Compiled sequence (top)	caccGAAACCAGAGGCAGCACTCTgttttAGAGCTAGAAATAGCAAGTTAAA ATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGgtgcTTC GTgCACCTAGAGTGCTGCCTCTGGTT
Compiled sequence (bottom)	aaaaAACCAGAGGCAGCACTCTAGGTGcACGAAGCACCGACTCGGTGCCA CTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGc tctaaaacAGAGTGCTGCCTCTGGTTTC

peg2_A	**Edits rs1799993 from protective to risk** Sequence (5’ → 3’)
Spacer sequence (top)	caccGAAACCAGAGGCAGCACTCTgtttt
Spacer sequence (bottom)	ctctaaaacAGAGTGCTGCCTCTGGTTTC
Extension oligo (top)	gtgcTAAAGTTCGTtCACCTAGAGTGCTGCCTC
Extension oligo (bottom)	aaaaGAGGCAGCACTCTAGGTGaACGAACTTTA
Compiled sequence (top)	caccGAAACCAGAGGCAGCACTCTgttttAGAGCTAGAAATAGCAAGTTAAA ATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGgtgcTA AAGTTCGTtCACCTAGAGTGCTGCCTC
Compiled sequence (bottom)	aaaaGAGGCAGCACTCTAGGTGaACGAACTTTAGCACCGACTCGGTGCCAC TTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGctc taaaacAGAGTGCTGCCTCTGGTTTC

peg3_A	Edits rs1799993 from protective to risk Sequence (5' → 3')
Spacer sequence (top)	caccTTCCATGTTGGTGAAACCAGgtttt
Spacer sequence (bottom)	ctctaaaacCTGGTTTCACCAACATGGAA
Extension oligo (top)	gtgcTTCGT**t**CACCTAGAGTGCTGCCTCTGGTTTCACCAACA
Extension oligo (bottom)	aaaaTGTTGGTGAAACCAGAGGCAGCACTCTAGGTG**a**ACGAA
Compiled sequence (top)	caccTTCCATGTTGGTGAAACCAGgttttAGAGCTAGAAATAGCAAGTTAAAA TAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGgtgcTTCG T**t**CACCTAGAGTGCTGCCTCTGGTTTCACCAACA
Compiled sequence (bottom)	aaaaTGTTGGTGAAACCAGAGGCAGCACTCTAGGTG**a**ACGAAGCACCGAC TCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGC TATTTCTAGctctaaaacCTGGTTTCACCAACATGGAA