Molecular techniques for therapeutic and diagnostic applications in Mucopolysaccharidosis IIIB and Gaucher disease

by

Chloe L. Christensen
B.Sc., University of Victoria, 2015

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Biology

We acknowledge with respect the Lekwungen peoples on whose traditional territory the university stands and the Songhees, Esquimalt and W̱SÁNEĆ peoples whose historical relationships with the land continue to this day.

Supervisory Committee

Molecular techniques for therapeutic and diagnostic applications in Mucopolysaccharidosis IIIB and Gaucher disease

by

Chloe L. Christensen
B.Sc., University of Victoria, 2015

Supervisory Committee

Dr. Francis Choy, Department of Biology
Supervisor

Dr. Raad Nashmi, Department of Biology
Departmental Member

Dr. Patrick Walter, Department of Biology
Departmental Member

Dr. Juan Ausio, Department of Biochemistry
Outside Member

Abstract

There is an unmet need to develop and test treatments for rare lysosomal disease (LD). Most LDs are present in childhood and do not currently have approved therapies. Rare diseases individually are uncommon but taken together account for a population prevalence of 3.5-5.9% worldwide. Due to their rarity, it often takes significant time and effort to diagnose rare diseases. New diagnostic tools, especially for early detection, will offer an advantage in avoiding this diagnostic odyssey. This dissertation is focused on investigating novel diagnostic and treatment methods *in vitro* for two neurodegenerative LDs: Gaucher disease (GD) and mucopolysaccharidosis IIIB (MPS IIIB). Mutations in *NAGLU* and *GBA1*, the genes that encode for lysosomal hydrolases required for degradation of heparan sulfate and glucocerebrosides, lead to the observed pathogenesis in MPS IIIB and GD, respectively. Since many LDs, including MPS IIIB and some forms of GD, are neurodegenerative, cell and gene-based therapeutic strategies are of significant interest. Therapeutics that offer some symptom mitigation in other LDs, such as enzyme replacement or substrate reduction therapies, do not offer appreciable disease mitigation in MPS IIIB or neurodegenerative GD.

Here, a novel compound heterozygous mutation, *NAGLU*$^{Y140C/R297X}$, that results in approximately 50% residual NAGLU protein and 0.6% NAGLU enzyme activity is reported in *NAGLU*. Furthermore, a RFLP and site-directed mutagenesis strategy was developed to identify the presence of the relatively common p.R297X mutation in patient cell samples, in addition to two other novel molecular assays for the detection of the p.E153K mutation in *NAGLU* and p.N370S mutation in *GBA1*. MPS IIIB and GD human skin fibroblasts were reprogrammed to iPSCs using non-integrating Sendai viral vectors with a reprogramming efficiency of 0.2% and 0.3%, respectively. Resulting iPS cell lines were confirmed as being pluripotent through a barrage

of analyses for markers of pluripotency and differentiation. Intriguingly, early passage MPS IIIB iPSCs were found to exhibit increased cell death and spontaneous differentiation to embryoid body-like structures, which was hypothesized to be caused by fibroblast growth factor 2 (FGF2) sequestration or degradation due to inherent heparan sulfate dysregulation. Supplemental FGF2 (100 ng/mL) was found to significantly increase confluency of MPS IIIB iPSCs after 48 hours ($n = 5$, $p \leq 0.05$) and persisting to 96 hrs ($n = 5$, $p \leq 0.05$), thus providing evidence for an important role of FGF2-heparan sulfate interactions in the maintenance of stem cell pluripotency. These findings highlight the importance of considering inherent disease pathology when developing disease models.

Three genome editing strategies, CRISPR-Cas9, base and prime editing, are addressed throughout this dissertation. Genome editing outcomes in *NAGLU* and *GBA1*, as well as a control gene, *HPRT1*, are reported in HEK293 cells, human skin fibroblasts, and induced pluripotent stem cells (iPSCs). Although CRISPR-HDR failed to yield mutation correction, base editing of the common p.N370S (c.1226 A>G) in GD skin fibroblasts using with 42% efficiency is reported. Base editing of *HPRT1* in HEK293 cells with an overall editing efficiency of $6 \pm 0.5\%$ ($n = 3$), but interestingly, when base editing at the centered nucleotide was analyzed, the editing efficiency increases to $27 \pm 4.3\%$ ($n = 3$). These findings align with other reports of a centered nucleotide preference for base editors and will help direct genome editing strategies in the future. This dissertation describes the first genome editing in *NAGLU*, and the first base editing in *GBA1*, and underscores the importance of optimizing genome editing strategies when targeting disease-causing mutations in patient-derived cells. The findings reported here will direct future genome editing strategies for developing cell and gene-based therapies for MPS IIIB and GD.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

%	percent
°C	degrees Celsius
ABE	adenine base editing
Akt	protein kinase B
ALP	alkaline phosphatase
APLS	alkaline phosphatase live staining
APOBEC1	apolipoprotein B mRNA editing enzyme, catalytic polypeptide 1
BBB	blood-brain barrier
bp	base pair
BSA	bovine serum albumin
Cas9	CRISPR-associated protein 9
CBE	cytidine base editing
cDNA	complementary DNA
cffDNA	cell-free fetal DNA
CHO	Chinese hamster ovary
CNS	central nervous system
CRISPR	clustered regularly interspaced short palindromic repeats
CRISPRa	CRISPR activation
CRISPRi	CRISPR interference
crRNA	CRISPR RNA
CVS	chorionic villus sampling
D-MEM	Dulbecco's modified eagle medium
DAPI	4′,6-diamidino-2-phenylindole
dCas9	dead Cas9
ddPCR	droplet digital PCR
dH_2O	distilled H_2O
DSB	double-stranded break
dsODN	double-stranded oligonucleotide
EB	embryoid body
ECM	extracellular matrix
EDTA	ethylenediaminetetraacetic acid
ERT	enzyme replacement therapy
ESC	embryonic stem cell
FB	fibroblast
FBS	fetal bovine serum
FGF2	fibroblast growth factor 2
FGFR	fibroblast growth factor receptor
GAG	glycosaminoglycan

GCase	glucocerebrosidase
GD	Gaucher disease
GD1	type 1 Gaucher disease
GD2	type 2 Gaucher disease
GD3	type 3 Gaucher disease
gDNA	genomic DNA
GFP	green fluorescent protein
GoF	gain of function
gRNA	guide RNA
GVHD	graft-versus-host disease
HDR	homology directed repair
hr	hour
HR	homologous recombination
HS	heparan sulfate
HSC	hematopoietic stem cell
HSCT	hematopoietic stem cell transplantation
HSPG	heparan sulfate proteoglycan
iMEF	irradiated mouse embryonic fibroblasts
indels	insertions/deletions
iPSC	induced pluripotent stem cells
kb	kilobase
kDa	kilodalton
LD	lysosomal disease
LoF	loss of function
M6P	mannose 6-phosphate
MAF	mutagenized forward primer
MAPK	mitogen-activated protein kinase
MAR	mutagenized reverse primer
MKOS	c-Myc, Klf4, Oct3/4, Sox2
MNP	motor neuron progenitor
MPC	multipotent progenitor cell
MPS	mucopolysaccharidosis
MPS IIIB	mucopolysaccharidosis IIIB
mRNA	messenger mRNA
MSD	multiple sulfatase deficiency
NAGLU	α-*N*-D-acetyl glucosaminidase
nCas	Cas nickase
nCas9	Cas9 nickase
NGS	next generation sequencing
NHEJ	non-homologous end joining

NIPD	non-invasive prenatal diagnosis
NIPT	non-invasive prenatal testing
NPC	neural progenitor cell
NSC	neural stem cell
PAM	protospacer adjacent motif
PBS	phosphate buffered saline
PCT	pharmacological chaperone therapy
PD	Parkinson's disease
pegRNA	prime editing gRNA
PI 3-K	phosphatidylinositol 3 kinase
PMP22	peripheral myelin promotor of 22 kDa
qPCR	quantitative PCR
RFLP	restriction fragment length polymorphism
RGEN-RFLP	RNA-guided endonuclease restriction fragment length polymorphism
RNP	ribonucleoprotein
RT	reverse transcriptase
SAP-C	saposin-C
SASD	sialic acid storage disease
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SeV	Sendai viral
sgRNA	single gRNA
SNP	single nucleotide polymorphism
SRT	substrate reduction therapy
SSEA	stage-specific embryonic antigen
ssODN	single-stranded oligonucleotide
STAT	signal transducer and activator of transcription
T7E1	T7 endonuclease I
TALEN	transcription activator-like effector nucleases
TAM	targeted AID-mediated mutagenesis
TAP	transcriptional activator protein
TH1-P	human monocytic cell line
tracrRNA	trans-activating crRNA
TSD	Tay-Sachs disease
UGI	uracil DNA glycosylase inhibitor
VTN-N	vitronectin
WT	wildtype
YFP	yellow fluorescent protein
ZFN	zinc finger nucleases

Acknowledgments

It is with unending appreciation that I acknowledge all of the important people in my life. You know who you are. A special acknowledgement to my mother and to my partner, Jonah, for always keeping me grounded. This work was made possible with generous grants from the Sanfilippo Children's Research Foundation, and the Rare Disease Foundation. Thank you also to the BC Regenerative Medicine Cluster and WORLD*Symposium* for providing a platform on which I could convey the progress of these projects and providing travel funds to allow me to share my research and make connections with future colleagues and collaborators. Thank you to my committee members, Dr. Raad Nashmi, Dr. Patrick Walter, and Dr. Juan Ausio, for their gracious support and suggestions during this process. Thank you to all the Choy lab members, especially Emma Wells-Durand, Rhea Ashmead, and Samuel Chu. Lastly, thank you to Dr. Francis Choy for his unwavering support during all of the ups and downs that are inevitably a part of research.

Dedication

Dedicated to all the children who have lived or are living with a rare disease.

May you be well.

1. Introduction

1.1. Lysosomal disorders

1.1.1. Overview

Lysosomes were first described by Christian de Duve in the 1950s, and for this discovery he was awarded the 1974 Nobel Prize for Physiology or Medicine (1955). Lysosomes are membrane-bound organelles that function as integral components of degradative pathways. Lysosomal hydrolases are responsible for the catabolism of macromolecules, resulting in the recycling of molecules used in numerous cellular processes. Lysosomes are at the crossroads of the autophagic, phagocytic and endocytic trafficking network, where they receive intra- and extracellular cargo for degradation (Xu and Ren, 2015; Glick *et al.*, 2010). A collection of more than 50 inherited disorders that ultimately result from increased lysosomal storage and the downstream impact of cargo turnover are referred to as lysosomal diseases (LDs) (Harlan *et al.*, 2016).

LDs were previously referred to as lysosomal storage diseases, but the acronym has recently changed to recognize emerging evidence that although the accumulation of macromolecules is a fundamental factor in LD pathology, there are numerous secondary and tertiary factors that add to the complexity of these disorders (Lysosomal Disease Network; Fecarotta *et al.*, 2018). Many LDs result from dysfunctional lysosomal hydrolases, including the mucopolysaccharidoses (MPSs), sphingolipidoses, and mucolipidoses, to name but a few. Other LDs result from deficiencies in other aspects of lysosome function, such as in sialic acid storage disease (SASD) where defects in a lysosomal membrane protein, sialin, prevents transport of acidic

monosaccharides across the lysosomal membrane (Mancini *et al.*, 1991; Tarailo-Graovac *et al.*, 2017). Multiple sulfatase deficiency (MSD), which results from mutations in the gene encoding for formylglycine-generating enzyme, the enzyme responsible for converting a cysteine to Cα-formylglycine, and thus renders sulfatases inactive (Cosma *et al.*, 2003). In this example, MSD is caused by an error in post-translational modification, indirectly impacting sulfatase function. Secondary and tertiary impacts of lysosomal storage include the activation of cellular pathways, impaired autophagy, and abnormal vesicular and plasma membrane trafficking. LDs range in prevalence from Gaucher disease (GD) with a prevalence of 1:57,000 to mucolipidosis I with an estimated prevalence of 1:4,200,000 live births (Meikle *et al.*, 1999). Collectively, these disorders are more common with an incidence of approximately 1:5000 live births (Vitner *et al.*, 2015). Additionally, certain LDs, such as Tay-Sachs disease (TSD) and GD, are much more common due to the founder effect in Ashkenazi Jewish populations (Landels *et al.*, 1991; Kolodny *et al.*, 1982). Due to their rarity, a number of these disorders are candidates for orphan drugs. As of 2019, there are 28 LDs for which FDA-approved compounds have been granted orphan drug status (Garbade *et al.*, 2020).

LDs are rare due to a number of reasons: (1) they are genetic disorders that typically follow autosomal recessive inheritance, (2) disease severity often prevents or delays progression to sexual maturity, or spontaneous abortion in individuals with attenuated LDs, thus reducing the likelihood of offspring of individuals with a LD homozygous genotype, and (3) a severe fetal LD phenotype can result in hydrops fetalis and fetal demise (Iaselli *et al.*, 2011; Zlotogora *et al.*, 1989; Stone and Sidransky, 1999). On the other hand, patients with late-onset or attenuated phenotypes may live long into adulthood. For example, patients with type 1 GD (GD1) are estimated to have a life

expectancy of only 9 years shorter than their reference population (Weinreb *et al.*, 2008). Additionally, carriers of LDs are by definition more frequent in the population than homozygotes. Given the abovementioned prevalence of GD, carrier frequency is calculated as approximately 1/120 individuals using the Hardy-Weinberg equation. Even though attenuation of LD phenotype results with only minimal lysosomal hydrolase activity (Oussoren *et al.*, 2013), and no discernable disease status is associated with LD mutation carriers, there is clear evidence for an association between perturbed lysosome function and aging-associated diseases (Ysselstein *et al.*, 2019). One pertinent example is the 4 to 6-fold increased likelihood for GD carriers and those with the non-neuronopathic subtype, GD1, developing Parkinson's disease (PD) later in life (Sidransky *et al.*, 2009; Bultron *et al.*, 2010). Recently, proteins affected in other LDs have been associated with Parkinsonism in patients or carriers, and/or observed α-synuclein accumulation (Ysselstein *et al.*, 2019; Winder-Rhodes *et al.*, 2011). Associated proteins include acid sphingomyelinase, galactosylceramidase, hexosamindase B, and α-*N*-acetyl glucosaminidase (NAGLU), which are enzymes rendered dysfunctional in Niemann-Pick disease type A and B, Krabbe disease, Sandhoff disease, and mucopolysaccharidosis IIIB (MPS IIIB), respectively (Ysselstein *et al.*, 2019; Galvan *et al.*, 2008; Chinen *et al.*, 2005; Keilani *et al.*, 2012; Suzuki *et al.*, 2003; Winder-Rhodes *et al.*, 2011). The association between *GBA1* mutations and an increased risk in developing PD is discussed in further detail in section 1.1.3.

1.1.2. Mucopolysaccharidosis IIIB

Overview

The mucopolysaccharidoses (MPSs) are inherited LDs that arise from an inability to degrade glycosaminoglycans (GAGs) that collectively have a prevalence of 1:25,000 live births, where MPS III is the most common within this group (Dorfman and Matalon, 1976; Baehner *et al.*, 2005; Meikle *et al.*, 1999). All seven MPS subtypes follow autosomal recessive inheritance, save MPS II which is X-linked (Hunter syndrome; D'Avanzo *et al.*, 2020). MPS IIIB results from mutations in the *NAGLU* gene (GCID: GC17P042535) which encodes for the lysosomal hydrolase NAGLU. NAGLU is required for hydrolysis of the α, 1→4 linkage between uronic acid and *N*-acetylglucosamine during the stepwise process of heparan sulfate (HS) degradation (Valstar *et al.*, 2008). Heparan sulfate proteoglycans (HSPGs) have numerous biologic activities at the cell surface and extracellular matrix (ECM) through interactions with growth factors, cytokines, and components of the ECM and cell surface themselves (Dreyfuss *et al.*, 2009). These interactions play roles in cell proliferation, differentiation, inflammation, and endocytosis, to name but a few (Dreyfuss *et al.*, 2009). In addition to cellular perturbations that result from lysosomal HS accumulation, dysregulated morphogen-receptor interactions ultimately impact signalling pathways (de Pasquale and Pavone, 2019). More than 150 pathogenic mutations, many of which are private, have been reported in *NAGLU* on chromosome 17q21.2 (Khorrami *et al.*, 2019). This gene is 8.3 kb, consisting of six exons and is proximal to the gene encoding 17ß-hydroxysteroid dehydrogenase 1, an enzyme involved in estradiol formation (Zhao *et al.*, 1996; Sasano *et al.*, 1996). Reported mutations in *NAGLU* include missense, nonsense, insertions/deletions (indels) and splice site mutations (Yogalingam and Hopwood, 2001). A spectrum of clinical disease

severity can arise, although genotype is not a reliable gauge of predicted disease severity (Yogalingam *et al.*, 2000; Valstar *et al.*, 2008).

Clinical manifestation of MPS IIIB

MPS IIIB (OMIM #252920) was first described in 1963 by Sylvester Sanfilippo, after whom the condition is sometimes referred to as Sanfilippo B syndrome (Sanfilippo *et al.*, 1963). MPS IIIB is a pediatric disorder, commonly manifesting in disease symptoms around 2-6 years of age (Delgadillo *et al.*, 2013; Kim *et al.*, 2016). Individuals with MPS IIIB have progressive cognitive decline, experience behavioural issues, hyperactivity, sleep deficits, and develop coarse facial features (Weber *et al.*, 1999; Yogalingam *et al.*, 2000; Lin *et al.*, 2018). MPS IIIB follows three phases of disease progression with respect to the neurodegenerative symptoms: (1) developmental delay, (2) behavioural issues, aggression, and reduced attention span and (3) seizures and the loss of motor skills (Cleary and Wraith, 1993). Skeletal dysplasia, coarse facial features and other somatic alterations in MPS III patients are not as severe as in other MPS subtypes, which has resulted in increased misdiagnoses (Cleary and Wraith, 1993). As is common with many rare disorders, patient's families often face a diagnostic odyssey as their child is misdiagnosed as having attention deficit/hyperactivity disorder or autism spectrum disorders (Wijburg *et al.*, 2013; Brady *et al.*, 2013). Significant clinical heterogeneity has been reported in MPS IIIB, and some patients even exhibit an attenuated form of the disease, although this is a rare occurrence (Weber *et al.*, 1999; Andria *et al.*, 1979; Zhao *et al.*, 1998).

Current therapeutics for MPS IIIB

The current standard of care for patients with MPS IIIB focuses on treating symptoms as they arise as there is no curative treatment (Shapiro *et al.*, 2019). Since the severity of MPS IIIB results from central nervous system (CNS) degeneration, novel treatments must be able to pass the blood-brain barrier (BBB), a system of capillaries acting as an anatomical barrier, to access the CNS. For a detailed review of the BBB and emerging methods to deliver therapeutics to the CNS, see Christensen *et al.* (2019). Lysosomal enzymes are post-translationally tagged with mannose-6-phosphate (M6P) in the Golgi apparatus and trafficked to the lysosomes (Sands and Davidson, 2006). A small proportion of tagged lysosomal enzymes are released into the extracellular fluid, where these tagged enzymes are then taken up by adjacent cells in a method known as cross-correction (Fratantoni *et al.*, 1968; Kaplan *et al.*, 1977; Sands and Davidson, 2006). In combination with a low residual enzyme threshold required to mitigate disease symptoms (1-15%), cross-correction offers an opportunity for systemic therapies that deliver functional lysosomal hydrolases to elicit a therapeutic effect (Andrade *et al.*, 2015; Oussoren *et al.*, 2013). Enzyme replacement therapy (ERT), substrate reduction therapy (SRT), and pharmacological chaperone therapy (PCT) are all techniques used to ultimately reduce toxic substrate accumulation and have been largely beneficial for many LDs; however, ERT, SRT and PCT do not mitigate CNS symptoms in MPS IIIB patients (Narita *et al.*, 2016; Patterson *et al.*, 2007; Machaczka *et al.*, 2012; Gaffke *et al.*, 2018).

Allogeneic hematopoietic stem cell therapy (HSCT) is the standard form of treatment for MPS IH (Parini *et al.*, 2017; de Ru *et al.*, 2011; Aldenhoven *et al.*, 2017), and recommended as a combination therapy for other LDs (Platt, 2014; Sands and Davidson, 2006), but is not recommended for MPS IIIB (Welling *et al.*, 2015) or GD (Somaraju *et al.*, 2017) due in part to a

lack of reduction of neurological symptoms (Sivakumur and Wraith, 1999; Vellodi *et al.*, 1992). Of all metabolic disorders, MPS IH is by far the most benefitted by HSCT and most studied, with over 500 MPS IH patients having received HSCT (Aldenhoven *et al.*, 2017). However, HSCT is not curative and only attenuates MPS IH (Aldenhoven *et al.*, 2017). Additionally, there is significant risk of graft-versus-host disease (GVHD) in patients receiving allogeneic HSCT (Mielcarek *et al.*, 2003). For now, the standard of care for MPS IIIB focuses on symptom treatment, including addressing ear, nose and throat infections that arise, providing corneal transplants, and surgical drainage of cerebrospinal fluid (Cleary and Wraith, 1993; National Institute of Neurological Disorders and Stroke).

1.1.3. Gaucher disease

Overview

GD is a rare LD characterized by the progressive accumulation of glucocerebrosides within lysosomes and was first described by Philippe Gaucher in 1882 (Mehta, 2006). This autosomal recessive disease results from mutations in *GBA1* (GCID: GC01M155234), which encodes the lysosomal hydrolase necessary for breaking down glucosylceramides to glucose and ceramide, glucocerebrosidase (glucosylceramidase, GCase). Accumulation of glucocerebrosides in macrophages is apparent upon histological interrogation (Bitton *et al.*, 2004). These “Gaucher cells” are enlarged with a wrinkled appearance, and are overloaded with iron (Bitton *et al.*, 2004). Patients with GD present with variable symptoms based on mutation severity and residual GCase activity, giving rise to three GD subtypes: GD1 (OMIM #230800), non-neuronopathic; type 2 (GD2; OMIM #230900), acute neuronopathic; and type 3 (GD3; OMIM #231000), sub-acute

neuronopathic. GD is the most commonly occurring LD and has an incidence of 1:450 and carrier frequency of ~9% in Ashkenazi Jewish populations (Zimran *et al.*, 1991). The most common *GBA1* mutation, resulting in 70% of GD, is c.1226 A>G (p.N370S) and is associated with non-neuronopathic GD (GD1) (Salvioli *et al.*, 2005). *GBA1* is located on chromosome 1q21, 16 kb downstream from a pseudogene, *GBAP1*, with 96% sequence similarity to *GBA1* (Martinez-Arias *et al.*, 2001; Ginns *et al.*, 1985). Over 500 mutations have been reported in *GBA1*, including missense, nonsense, indels, splicing variants, complex rearrangements and recombinant mutations (Romero *et al.*, 2019; Bhutada *et al.*, 2018; Mitsui *et al.*, 2015; Dimitriou *et al.*, 2020). Many GD1 or GD3 patients have the common p.L444P (c.1448 T>C) missense mutation in either homozygosity or in compound heterozygosity with another mild mutation (Wan *et al.*, 2006). c.1448 T>C has been found to account for 53.5% of sequenced GD1 and GD3 alleles (Wan *et al.*, 2006). Other common GD pathogenic mutations include the recombinant allele c.1448 T>C/c.1483 T>G/c.1497 G>C (Rec*Nci*I) and the IVS2 +1 splicing variant (Saleem *et al.*, 2017).

Clinical manifestations of GD

Patients with the most severe form of GD (GD2) succumb to the disease in infancy due to the associated progressive neurodegeneration; however, patients with non-neuronopathic (GD1) or subacute neuronopathic (GD3) forms of the disease live into adulthood. Patients with GD experience hepatosplenomegaly, osteopenia, osteonecrosis, and cytopenia due to the accumulation of Gaucher cells (Guggenbuhl *et al.*, 2008; Stirnemann *et al.*, 2017). As previously mentioned, the life expectancy of GD1 patients is only slightly less than normal, with estimates of life expectancy of about 68 years (Weinreb *et al.*, 2008; Burrow *et al.*, 2011). GD3 is much more clinically

heterogenous, with life expectancies ranging from the 10th to the 40th decade (Burrow *et al.*, 2011). Mutations in *PSAP*, which encodes for the pro-peptide of a protein required in glucocerebroside degradation, saposin-C (SAP-C), has also been found to result in a GD phenotype (Tylki-Szymanska *et al.*, 2007). SAP-C variants have been hypothesized as a modifier of clinical GD severity, and may partly explain disease heterogeneity (Tamargo *et al.*, 2012). Clinical case studies have shown a marked concurrence of GD and the second most prevalent neurodegenerative disorder, PD, in patients with GD and heterozygous carriers for *GBA1* mutations (Sidransky *et al.*, 2009).

An association between Parkinson's disease and GD

PD, which affects ~0.2% of Canadians, was historically classified as an idiopathic disease (Statistics Canada, 2001). However, a genetic basis for PD has been more recently elucidated in ~5% of cases (Bekris *et al.*, 2010; Fernandes *et al.*, 2016). Patients with PD exhibit neurological deficits affecting the motor system, resulting in rigidity, slowed movement, higher order processing issues, and dementia (Itokawa *et al.*, 2006). Motor symptoms result from dopaminergic neuron death within the substantia nigra due to the presence of Lewy bodies, a main component of which is the protein product from the *SNCA* gene, α-synuclein (Spillantini *et al.*, 1997). Mutations in *PINK1*, *DJ1*, *ATP13A2*, and *PRKN* are causative for PD, and mutations in *SNCA*, *LRRK2*, and *GBA1* are a risk factor for the development of PD, where *GBA1* mutations are the most common PD risk factor (Rocha *et al.*, 2015; Aflaki *et al.*, 2017). Despite numerous ongoing investigations, a molecular pathway to explain the etiology of *GBA1*-associated PD is still lacking (Fernandes *et al.*, 2016). Early clinical case studies report an increased association of patients with

GD1 developing PD later in life (Neudorfer *et al.*, 1996; Machaczka *et al.*, 1999; Tayebi *et al.*, 2001; Varkonyi *et al.*, 2002; Bembi *et al.*, 2003). Upon large-scale genetic screening studies, patients with GD or carriers for *GBA1* mutations have been shown to have a 4 to 6-fold increased risk for developing PD (Sidransky *et al.*, 2009; Bultron *et al.*, 2010). Based on the current body of evidence, gain of function (GoF) and loss of function (LoF) hypotheses for *GBA1*-associated PD disease progression have been proposed (Campbell and Choy, 2012; Aflaki *et al.*, 2017). The GoF hypothesis suggests that mutant GCase aggregates with an ubiquitin E3 ligase, parkin, impairing endoplasmic reticulum-associated degradation, which negatively impacts α-synuclein turnover and leads to the sequestration of insoluble α-synuclein oligomers, and ultimately dopaminergic neuron death. Interestingly, null mutations show an even greater risk for PD with an earlier age of onset over mild to severe *GBA1* mutations, which conflicts with the GoF hypothesis (Campbell and Choy, 2012; Swan and Saunders-Pullman, 2013; Ron *et al.*, 2010).

The LoF hypothesis suggests that a lack of GCase activity leads to the accumulation of glucocerebrosides resulting in lipid homeostasis dysregulation, which in turn impacts α-synuclein processing (Campbell and Choy, 2012; Aflaki *et al.*, 2017). The LoF hypothesis, however, does not account for *GBA1* mutations exhibiting incomplete penetrance for PD or for *GBA1* heterozygous individuals showing an increased risk for PD (Aflaki *et al.*, 2017). Thus far, study designs have included the use of GCase inhibitors, comparisons of multiple wildtype and *GBA1* mutant cell lines and mouse models, and analyses of the molecular interactions between α-synuclein and GCase, but there have not yet been investigations into the common homozygous c.1226 A>G mutation (Mazzulli *et al.*, 2011; Sun *et al.*, 2015; Manning-Bog *et al.*, 2009; Xu *et al.*, 2011).

Current therapeutics in GD

The current standard of care for GD1 is ERT or SRT for reducing a patient's visceral and hematological symptoms (Bennett and Mohan, 2013; Roshan Lal and Sidransky, 2017). However, these treatments must be repeated regularly and cost upwards of $200,000 CDN/year (van Dussen *et al.*, 2014). ERT and SRT are not recommended for neuronopathic forms of GD (Roshan Lal and Sidransky, 2017; Schiffmann *et al.*, 2008). Investigations into other viable treatment options, including PCT and adeno-associated viral therapies, are ongoing, but are not without their limitations (Hong *et al.*, 2004).

1.2. Genome editing strategies

1.2.1. Overview and history

Recent progress made in the development and stratification of genome editing technologies has made it possible to modify eukaryotic genomes with precision (Jinek *et al.*, 2012). Genome editing systems have been used not only for modification of DNA sequence, but also gene expression modulation, editing of the epigenome, and fluorescent tagging, to name but a few (Chen *et al.*, 2018; Kang *et al.*, 2019; Laufer and Singh, 2015). Significant interest has gone into the development of genome editing systems for *in*- or *ex vivo* DNA modification for the treatment of human disease. Early iterations of genome editing systems include zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and homing endonucleases (Gaj *et al.*, 2013; Belfort and Bonocora, 2014). These tools are used to stimulate homologous recombination (HR) at a site of interest by introducing a double-stranded break (DSB) at a predictable location

(Bishop and Schiestl, 2000; Porteus and Carroll, 2005). An exogenous DNA template can be incorporated at the site of the break via homology directed repair (HDR), a process which was first reported in yeast (Hinnen *et al.*, 1978; Orr-Weaver *et al.*, 1981). Without the presence of a DNA template, the break is repaired haphazardly through the process of non-homologous end joining (NHEJ), which introduces indels at the site of the break (Pastwa and Blasiak, 2003).

ZFNs, originally termed chimeric restriction endonucleases, utilize FokI restriction endonucleases and modifiable zinc finger DNA binding domains that recognize trinucleotides to target specific genomic locations and cleave DNA (Kim *et al.*, 1996; Chandrasegaran and Carroll, 2016). Similarly, TALENs make use of FokI endonuclease domains, but use individual transcription activator-like effectors that each recognize a single DNA nucleotide, making TALENs easier to engineer and more customizable than ZFNs (Joung and Sander, 2013). There are significant issues with these systems, including increased cytotoxicity of ZFNs and large cargo size of TALENs, rendering these systems less desirable than more recent genome editing iterations (Mussolino *et al.*, 2011; Gupta and Musunuru, 2014). Clustered regularly interspaced short palindromic repeats (CRISPR)-Cas system was first referenced as a peculiarity in Ishino *et al.*'s 1987 report on the *iap* sequence in *Escherichia coli*. They noted regular spacing of repeat sequences in the 3' region flanking *iap*, and that a biological role for these sequences was not yet known (Ishino *et al.*, 1987). Since this report, CRISPR-Cas systems have been found to be responsible for protection from bacteriophages, acting as an acquired bacterial immune system in 45% of bacteria (Rath *et al.*, 2015).

The CRISPR-Cas acquired bacterial immune system has three phases of immunity: (1) adaptation, (2) expression, (3) and interference (Rath *et al.*, 2015). During adaptation, a bacteriophage infects a bacterium, and the Cas1-Cas2 complex recognize a protospacer adjacent motif (PAM) within the introduced viral DNA, cleave the DNA and incorporate the sequence into the bacterium's genome within the CRISPR array (Rath *et al.*, 2015). These incorporated spacer sequences are separated by identical repeat sequences (Rath *et al.*, 2015). Upon a subsequent infection, the CRISPR array is expressed and 5'-3' spacer RNA sequences complex with a Cas nuclease, such as Cas9 in type II CRISPR-Cas systems, leading to target binding through spacer complementarity and target degradation catalyzed by the Cas endonuclease (Rath *et al.*, 2015). The RNA that complexes with Cas endonucleases are comprised of a 5' CRISPR RNA (crRNA) complementary to the invading bacteriophage DNA and a 3' trans-activating crRNA (tracrRNA) that interacts with Cas (Jinek *et al.*, 2012).

This CRISPR-Cas9 machinery was exploited to use a modified crRNA designed to target a genomic site of interest, and for this demonstrated application of CRISPR-Cas systems as putative genome editing tools, Emmanuelle Charpentier and Jennifer Doudna were awarded the 2020 Nobel Prize in Chemistry (Jinek *et al.*, 2012; Ledford and Callaway, 2020). Since this system was demonstrated as a genome editing tool in 2012, CRISPR-Cas9 has been adapted for numerous genome editing applications; however, hereafter, I will focus on genome editing applications for the treatment of human genetic disease. Genome editing of the ß-globin gene, *HBB*, is one early example of adaptations of this programmable CRISPR-Cas9 system to yield editing outcomes with putative clinical applications for Sickle cell anemia (Xie *et al.*, 2014). In this example, a guide RNA (gRNA; crRNA and tracrRNA) targeting intron 1 of *HBB* is used for DSB creation (Xie *et*

al., 2014). HR using a DNA construct that encompasses two mutations proximal to exons 1 and 2, plus 0.5 kb of 3' and 5' flanking WT sequence, allows for the insertion of a *piggyBac* construct complete with selection markers, *puroΔtk* and *Neo*, which are removed via transient expression of transposase (Xie *et al.*, 2014). Xie *et al.* performed CRISPR-Cas9 genome editing in human induced pluripotent stem cells (iPSCs) which were then differentiated to erythrocytes, thus demonstrating proof-of-concept somatic cell editing of *HBB* (Xie *et al.*, 2014). Human germline genome editing has also been demonstrated in addition to somatic editing. Ma *et al.* corrected a heterozygous *MYBPC3* mutation, which leads to inherited autosomal dominant hypertrophic cardiomyopathy, in human preimplantation embryos (2017). This study suggests a putative role of genome editing in conjunction with *in vitro* fertilization to increase the chances of obtaining healthy embryos for transfer when a pathogenic paternal allele is present (Ma *et al.*, 2017).

Thus far, CRISPR HDR has been discussed as a suitable method for introducing sequence modifications at a site of interest. However, use of an exogenous HDR template is not without its drawbacks. HDR occurs far less frequently than NHEJ, with HDR typically observed in less than 10% of transfected cells (Devkota *et al.*, 2018; Anzalone *et al.*, 2019). Therefore, techniques to increase the efficiency of HDR are of interest to the field of genome editing. HDR normally occurs during S and G2 phase of the cell cycle, thus reversible cell cycle synchronization using cyclin-dependent kinase inhibitors, such as nocodazole, have been shown to significantly increase HDR efficiency (Yiangou *et al.*, 2019; Lin *et al.*, 2014). Cells selectively edited in G2/M phase compared to G1 phase present a 6-fold increase in targeting efficiency (Yang *et al.*, 2016). Lin *et al.* studied the effect of nocodazole, as well as lovastatin, mimosine, aphidocolin, thymidine, and hydroxyurea, on overall HDR efficiency in HEK293T cells and found HDR to be increased up to

38% from 10% (2014). Additionally, physical properties of HDR templates have been shown to greatly impact the editing efficiency (Lin *et al.*, 2014; Richardson *et al.*, 2016; Li *et al.*, 2014). In addition to studying the impact of cell cycle synchronizers on HDR efficiency, Lin *et al.* looked at the role of HDR templates with different length homology arms (2014). They were also interested in the impact of using a double-stranded oligonucleotide (dsODN) compared to a single-stranded oligonucleotide (ssODN) (Lin *et al.*, 2014). They found that symmetrical homology arms of more than 60 nucleotides show HDR efficiencies of more than ~19%, in comparison to <10% when only 30 nucleotide homology arms were used (Lin *et al.*, 2014). They also found that dsODN with 90 nucleotide homology arms yield 7% HDR efficiency compared to 20-23% for their ssODN counterpart, indicating that ssODN templates are more efficient at generating HDR (Lin *et al.*, 2014). Others have looked at the impact of symmetrical vs. asymmetrical HDR templates, in addition to optimizing donor templates to fit existing models of strand invasion during the HDR process (Richardson *et al.*, 2016). These investigators found that ssODNs with 36 nucleotides PAM distal and 91 nucleotides PAM proximal, and complementary to the non-target strand, generated the highest HDR efficiencies (57 ± 5%) when compared to 19 other physical HDR template designs in both HEK293 and K562 cells and at 6 different loci (Richardson *et al.*, 2016; Figure 1.1).

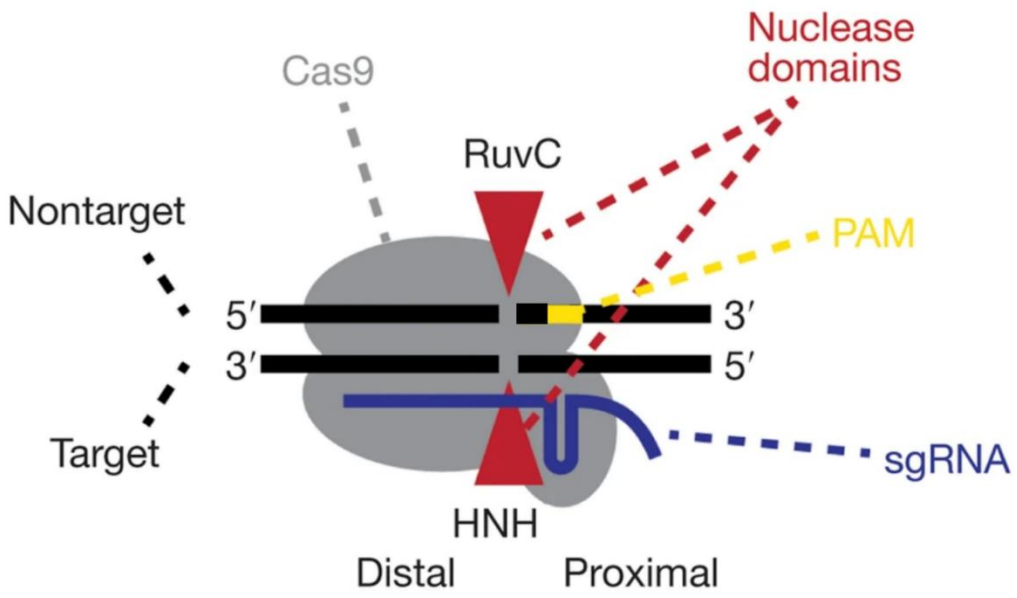


Figure 1.1 Components of the CRISPR-Cas9 genome editing system and relative target orientation, strand sense, and location of nuclease domains. Cas9 (grey) creates a DSB at a genomic site of interest as directed by a single gRNA (sgRNA, blue). Cas9 has two catalytic domains: RuvC and HNH. HNH cleaves the target strand and RuvC cleaves the non-target strand, with respect to the sgRNA target, three nucleotides upstream from the protospacer adjacent motif (PAM, yellow), which presents in the form of 5'-NGG-3', where N indicates any nucleotide. PAM proximal and distal regions are defined (modified from Richardson *et al.*, 2016).

Ensuring accurate spatial and temporal delivery of the HDR template so that it is available for incorporation at the Cas9 cut site has also been hypothesized to increase HDR efficiency (Savic *et al.*, 2018). Savic *et al.* show a 10 to 24-fold increase in HDR efficiency in HEK293T cells by covalently linking the ssODN HDR template to Cas9 (2018). They show that this increase in HDR efficiency is due to increased nuclear localization of the HDR template in comparison to un-coupled HDR template/Cas9 constructs and predict this will be a useful tool for future clinical

applications of the CRISPR-Cas9 system (Savic *et al.*, 2018). In addition to increasing editing efficiency with the CRISPR-Cas9 system, off-target editing events must be addressed. Off-target events can occur when nucleotide mismatches are tolerated, leading to Cas9 cleavage and indels at unintended sites within the genome (Wu *et al.*, 2014). Some evidence for a mismatch intolerance for nucleotides nearest to the PAM site (termed the 'seed' region, 3' end of the gRNA) has emerged, whereas the 5' gRNA is more tolerant to mismatches (Wu *et al.*, 2014). *In silico* tools have been designed to predict the likelihood and putative impact of off-targeting and choose gRNAs to minimize this effect, but these tools do not always recapitulate editing outcomes *in vivo* and *in vitro* (Mali *et al.*, 2013; Hsu *et al.*, 2013; Fu *et al.*, 2013; Cho *et al.*, 2014). It is recommended that these sites are sequenced post-editing to rule-out off-target impact (Zischewski *et al.*, 2017). Understanding and minimizing the impact of off-target editing will be essential in translating genome editing techniques to the clinic.

1.2.2. Genome editing for LDs

Genome editing is a particularly useful technique for creating models of LDs and also as a putative therapeutic strategy for a number of reasons: (1) LDs are monogenic, (2) some LDs lack adequate *in vivo* disease models, (3) many LDs lack suitable treatments that increase quality of life for patients and caregivers, (4) only small increases in enzyme levels are needed to offer therapeutic benefit, and (5) functional lysosomal enzymes are trafficked to neighbouring cells resulting in cross-correction (Bunge *et al.*, 1998; Fratantoni *et al.*, 1968; Kaplan *et al.*, 1977; Sands and Davidson, 2006). CRISPR-based genome editing strategies and their utility for modeling and

treating LDs has been reviewed at length in Christensen and Choy (2017), and Christensen *et al.* (2019). Recent uses of genome editing for modeling and treating LDs will be discussed here.

The majority of genome editing for LDs reported in the literature focuses on modeling LDs *in vitro* and *in vivo*, typically using either a knock-out strategy whereby NHEJ is utilized to generate indels in the gene of interest, or in some cases introducing a specific mutation to recapitulate the disease in a murine model (Pavan *et al.*, 2020; Huang *et al.*, 2020). CRISPR-Cas9 has so far been used to create *in vitro* models of lysosomal acid lipase deficiency, Fabry disease, GD, GM1 gangliosidosis, Niemann-Pick type C disease and Lowe syndrome in a range of disease-specific and transformed cell types (Zhang *et al.*, 2017; Song *et al.*, 2016; Pavan *et al.*, 2020; Latour *et al.*, 2019; Erwood *et al.*, 2019; Barnes *et al.*, 2018). Pavan *et al.* utilized CRISPR-Cas9 to knock-out *GBA1* in the human monocytic cell line (THP-1) and the glioblastoma U87 cell line, demonstrating reduced GCase activity, accumulation of LysoGL1, endoplasmic reticulum stress, increased cytokine levels, and an increased accumulation of α-synuclein (2020). However, they do not look specifically at the interaction between mutant GCase and α-synuclein in this particular study (Pavan *et al.*, 2020). As of this writing, there are no studies that look at the impact of introducing known human *GBA1* mutations to isogenic control cells. Murine models of Morquio syndrome type B, GM1 gangliosidosis, and infantile-onset Pompe disease, and an ovine model of neuronal ceroid lipofuscinosis have all recently been generated using CRISPR-Cas9 and demonstrate expected disease characteristics (Przybilla *et al.*, 2019; Eaton *et al.*, 2019; Huang *et al.*, 2020). Recently, CRISPR-based methods for correcting LD-causing mutations *in vitro* have also been demonstrated (Allende *et al.*, 2018; Valetdinova *et al.*, 2019). This includes the correction the c.840 C>T mutation in *SMN2*, a mutation that causes spinal muscular atrophy due

to decreasing levels of SMN protein (Valetdinova *et al.*, 2019). Valetdinova *et al.* used CRISPR-Cas9 and CRISPR-Cpf1 alongside ssODN HDR templates to correct this mutation in skin fibroblasts (FBs), iPSCs, and motor neuron progenitors (MNPs), showing greater editing efficiency using the CRISPR-Cas9 system in FBs and iPSCs over MNPs (Valetdinova *et al.*, 2019). One interesting strategy using CRISPR-Cas9 and two gRNAs allows for the deletion of a splice site mutation in *GLA*, which encodes α-galactosidase, the enzyme non-functional in Fabry disease (Chang *et al.*, 2017). Chang *et al.* found that using this approach in patient FBs significantly increased α-galactosidase activity and reduced the accumulation globotriaosylceramide in these cells (2017).

Similar methods utilizing CRISPR-Cas9 have restored enzyme activity in a murine model for MPS I and knocked out the promotor for the gene encoding the peripheral myelin protein of 22 kDa (PMP22), which is overexpressed in Charcot-Marie-Tooth disease type 1A and results in demyelination (Schuh *et al.*, 2018; Lee *et al.*, 2020). Significant interest has gone instead into the introduction of the gene implicated in certain LDs into a safe harbour locus, such as CCR5 or the albumin locus (Scharenberg *et al.*, 2020; Ou *et al.*, 2019; Ou *et al.*, 2020). The safe harbour approach with ZFNs has even led to clinical trials as of late for the treatment of MPS I and MPS II, Hurler and Hunter syndromes, respectively (NCT02702115; NCT03041324). However, these techniques are expected to only be useful for attenuated forms of MPS I and MPS II without severe neurological involvement (Poletto *et al.*, 2020). Although interim results of these studies show that the treatment is well tolerated, increased plasma enzyme levels are limited (NCT02702115; NCT03041324; Poletto *et al.*, 2020). Scharenberg *et al.* have recently used CRISPR-Cas9 to knock-in *GBA1* expression cassettes at the CCR5 safe harbour locus in HSCs under a lineage-

specific promotor and show increased enzyme activity in macrophages (2020). Safe harbour knock-in strategies do not circumvent the secondary issues that arise from enzyme GoF, such as the role of GCase and α-synuclein interactions discussed in the previous section. The aforementioned progress in genome editing methodologies is exciting for the area of LD research, but existing issues in CRISPR-Cas9 or ZFN techniques such as off-target effects, activation of p53 and increased apoptosis associated with DSBs, and *in vivo* construct delivery logistics must be addressed (Haapaniemi *et al.*, 2018; Lino *et al.*, 2018; Zischewski *et al.*, 2017).

1.2.3. Base and prime editing strategies

CRISPR-Cas9 is exceedingly effective for creating gene knock-outs by targeting an early exon in a gene of interest, resulting in indels and increasing the likelihood of a non-functional gene product (Jin *et al.*, 2020; Maruyama *et al.*, 2015). However, there are inherent issues with the CRISPR HDR system, including: (1) HDR occurs with low incidence, typically <10% or much lower, (2) DSBs have been shown to result in cellular apoptosis due to the activation of p53, (3) CRISPR HDR requires a separate template molecule, thus introducing spatial and temporal variables, and (4) HDR has been shown to occur in only certain stages of the cell cycle (Anzalone *et al.*, 2019; Maruyama *et al.*, 2015; Mali *et al.*, 2013; Chang *et al.*, 2017; Yeh *et al.*, 2018; Lin *et al.*, 2014; Heyer *et al.*, 2010; Haapaniemi *et al.*, 2018). To overcome these inherent issues associated with CRISPR HDR, recent alternative strategies have been developed (Komor *et al.*, 2016; Gaudelli *et al.*, 2017; Koblan *et al.*, 2019; Anzalone *et al.*, 2019).

Base editing

The first iteration of a programmable base editing system was proposed by Komor *et al.* primarily to address the issues associated with DSBs (2016). They engineered a genome editing system wherein a single nucleotide could be precisely targeted using a sequence-specific gRNA and nCas9, attached to a cytosine deaminase moiety capable of deaminating cytosine to uracil, with precision offered by CRISPR HDR but without requiring a DSB or a spatial and temporal HDR template (2016; Figure 1.2). They utilized the rat apolipoprotein B mRNA editing enzyme, catalytic polypeptide 1 (APOBEC1), a cytidine deaminase (Saraconi *et al.*, 2014). APOBEC1 has been found to deaminate a specific cytosine in the mRNA transcript of the human apolipoprotein B, and *APOBEC1* is highly expressed in human esophageal adenocarcinomas and results in hepatocellular carcinomas in rodents when constitutively expressed (Saraconi *et al.*, 2014; Blanc *et al.*, 2010; Yamanaka *et al.*, 1995). gRNA binding leads to the formation of an R-loop, which in conjunction with spatial restrictions imposed by nCas9-deaminase linker length, a predictable five-nucleotide base editing window located 13-17 nucleotides 5’ of the PAM is created (Komor *et al.*, 2016). Any cytosine within this window is a putative target for deamination; however, it has recently been noted that the base located at the center of this window (position 15 relative to the PAM) is more efficiently edited than other locations within the window (Anzalone *et al.*, 2020). Komor *et al.* found that upon DNA replication or repair, cytosine was not retained but instead reverted to uracil by base excision repair (2016). This group sought to mitigate this issue by introducing uracil DNA glycosylase inhibitors (UGIs), and also induce mismatch repair by using a nCas9 that nicks the non-edited strand, favouring conversion of the original cytosine to thymine (Komor *et al.*, 2016). This cytosine base editing (CBE) system is promising for editing pathogenic

mutations in human cells, for 37% of DNA sequences they analysed post-editing showed the intended cytosine to thymine conversion and only 1.1% indels in HEK293T, a substantial improvement from editing efficiencies and indels generated with CRISPR HDR (Komor *et al.*, 2016).

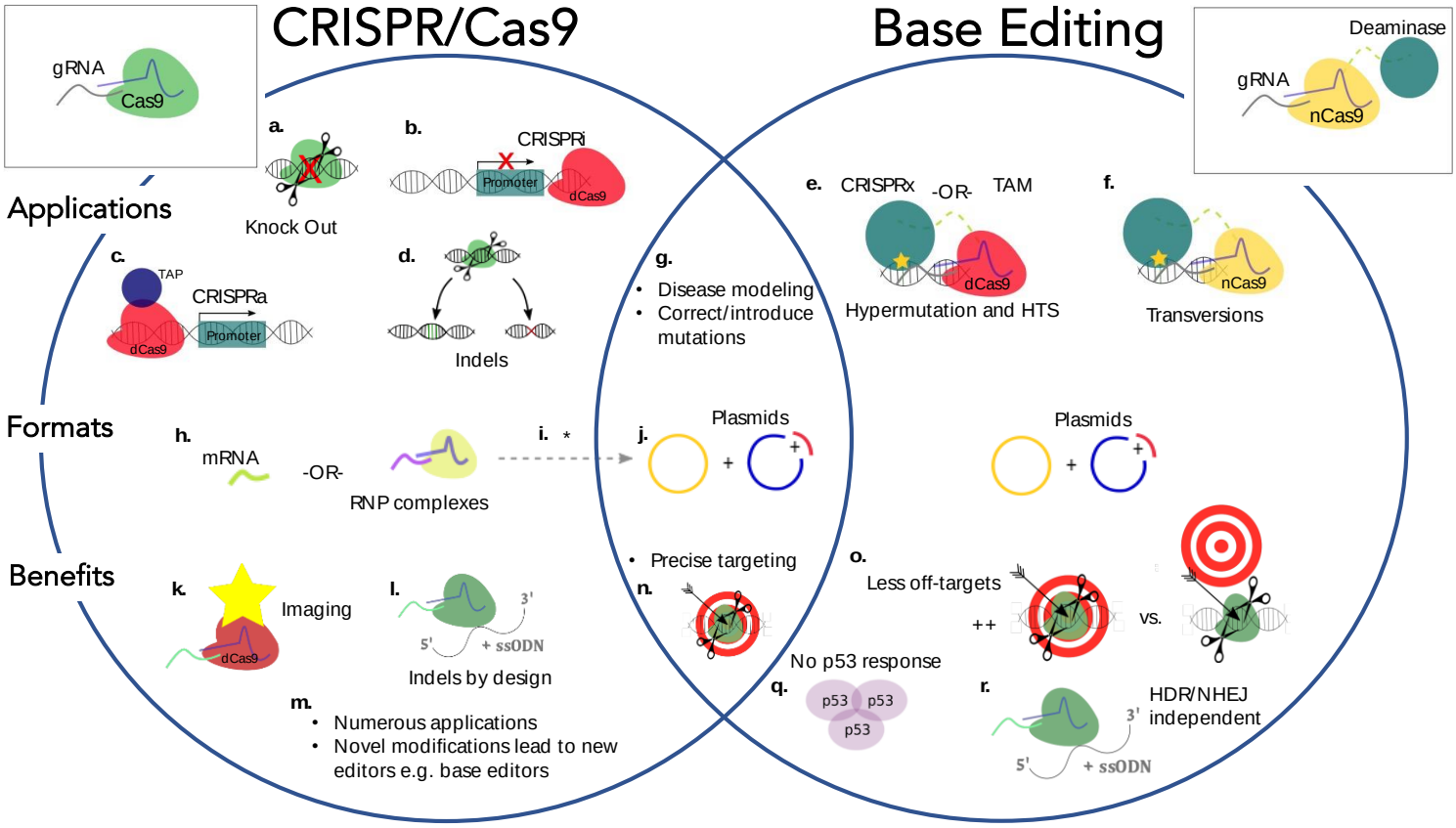


Figure 1.2. A comparison of the CRISPR-Cas9 and base editing systems on the basis of applications, formats, and benefits. Applications for CRISPR-Cas9 and base editing (a.–g.). CRISPR-Cas9 genome editing can be used to create DSBs, followed by the employment of NHEJ for gene knock-outs through the introduction of indels (a.) or precise large insertions or deletions through the use of homology HDR and a correction template (d.). Using a dead Cas9 (dCas9), CRISPR interference (CRISPRi) (Larson *et al.*, 2013) or CRISPR activation (CRISPRa) (Chavez *et al.*, 2016) to introduce a transcriptional activator protein (TAP) for recognition by RNA polymerase and activation at a promoter site (c.), or to bind downstream of a promoter thereby blocking RNA polymerase (b.), respectively. Base editing applications include CRISPRx or targeted AID-mediated mutagenesis (TAM) (Hess *et al.*, 2016) (e.), which utilizes a dCas9 and deaminase moiety for hypermutation applications and high throughput screening. Precise transversion mutations are realized through cytidine and adenine base editors (Komor *et al.*, 2016; Gaudelli *et al.*, 2017) (f.) Both CRISPR-Cas9 and base editing systems are capable of assisting in the creation of cell lines and mouse models for disease modeling and can be used to introduce or correct mutations (g.) Formats for CRISPR-Cas9 and base editing (h.–j.). Formats for CRISPR-Cas9 genome editing include ribonucleoprotein (RNP) complexes, mRNA, or plasmids encoding the necessary components (h.–j.). Base editors are currently available as plasmids but may become available in mRNA or RNP complex formats in the future*. Benefits for CRISPR-Cas9 and base editing (k.–r.). Additional benefits of CRISPR-Cas9 genome editing include gene imaging (Duan *et al.*, 2018) (k.) and the creation of precise

insertions (i.), among numerous other previously mentioned applications. Modifications to the CRISPR-Cas9 system continue to bring about novel methods, such as the base editing system (m.). Base editors show less off-targets (Gaudelli *et al.*, 2017) (o.), a diminished p53 response (Haapaniemi *et al.*, 2018) (q.), and are independent of cellular repair systems (Gaudelli *et al.*, 2017) (r.), HDR and NHEJ, in comparison to its CRISPR-Cas9 counterpart. Both systems are highly precise with regards to on-targeting (n.). (Modified from Christensen *et al.*, 2019).

Complementing the aforementioned CBE system, Gaudelli *et al.* set out to engineer an adenine base editing (ABE) system capable of editing cytosine to thymine transition mutations (2017). Due to spontaneous deamination of methylated cytosine, cytosine to thymine transition mutations account for the majority of known human pathogenic mutations (Antonarakis *et al.*, 2000). By swapping the APOBEC1 deaminase of the CBE system with a modified *E. coli* TadA adenine deaminase, Gaudelli *et al.* generated ABEs capable of converting adenine to guanine with an efficiency of 50% in human cells (2017). Recently, base editors have been used to create point mutations and correct pathogenic mutations in human tripronuclear zygotes, hematopoietic stem cells (HSCs), and in mice, to name but a few (Zhou *et al.*, 2017; Levy *et al.*, 2020; Zeng *et al.*, 2020). Taken together, ABEs and CBEs are in principle capable of targeting 48% of known pathogenic human single nucleotide polymorphisms (SNPs) and 30% of all known pathogenic mutations (Gaudelli *et al.*, 2017; Anzalone *et al.*, 2019). Furthermore, 57% of known pathogenic mutations in the MPSs are targetable using these systems (Figure 1.3; Christensen *et al.*, 2019). The applicability of base editing systems for the MPSs are discussed at length in Christensen *et al.* and will not be discussed here (2019).

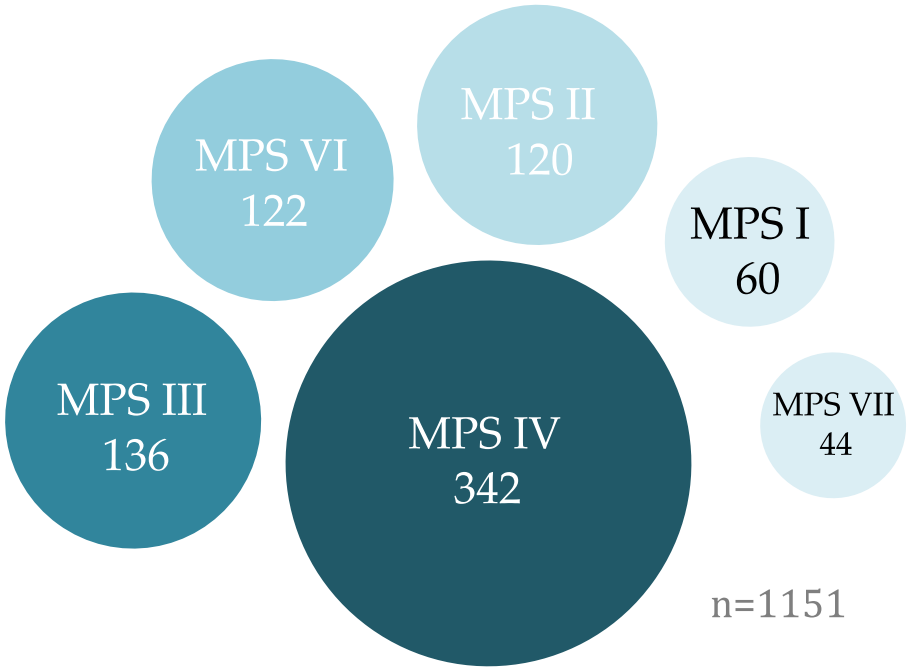


Figure 1.3. Graphical representation of mucopolysaccharidosis (MPS) mutations as targets for cytosine base editors (CBEs) or adenine base editors (ABEs). A small majority (57%) of known MPS disease-causing mutations (n = 1151) are currently targetable with existing ABE and CBE systems (ClinVar, accessed on 3 December 2018; Christensen *et al.*, 2019).

Prime editing

Despite the aforementioned improvements of base editing systems over CRISPR-Cas9, base editors are incapable of editing transversion mutations, introducing or removing indels, and are complicated by bystander base editing (Anzalone *et al.*, 2020). Anzalone *et al.* developed a hybrid genome editing system, termed prime editing, that can function to introduce a wide variety of mutations or edits, including transversions, transitions, and indels without HDR or the introduction of DSBs. This system relies on a nCas9 attached to a reverse transcriptase (RT) through a linker, and possesses an extended gRNA molecule, termed a prime editing gRNA

(pegRNA), which acts as a template for RT toward the 3' end of the pegRNA molecule (Anzalone *et al.*, 2019). The terminal 3' end of the pegRNA contains a primer binding site, which is complementary to the ~11 nucleotides 5' of the nCas9 nick site (Anzalone *et al.*, 2019). Through binding of the 3' pegRNA to the excised DNA at the nCas9 nick site, this creates a primer at which RT can synthesize complementary DNA (cDNA) from the pegRNA template, thus incorporating precise alterations in the DNA sequence (Anzalone *et al.*, 2019). The versatility offered by the prime editing system results in a stark increase in possible human pathogenic mutations as targets for editing, from 30% with base editors to 89% with prime editors (Gaudelli *et al.*, 2017; Anzalone *et al.*, 2019). Using this technique, Anzalone *et al.* demonstrated numerous edits in human cells, including a 3-nucleotide insertion (CTT) at the *HEK3* locus, the correction of a transversion in *HBB*, and the removal of a four-nucleotide duplication in *HEXA*, the gene implicated in TSD (2019). In addition to their versatility, prime editors, like base editors, are typically more efficient than CRISPR-HDR (Gaudelli *et al.*, 2017; Anzalone *et al.*, 2019). However, base editors are currently recommended over prime editors when a suitable base editing window free of bystander bases exists, since base editing typically results in fewer indels (Anzalone *et al.*, 2019).

Recently, Surun *et al.* demonstrated prime editing in human iPSCs by converting a tyrosine in a green fluorescent protein (GFP) construct to a tryptophan, effectively converting GFP to cyan fluorescent protein (2020). By modifying the length of the primer binding site in the pegRNA between 11-15 nucleotides, they found that an increased primer binding site length greatly decreased the editing efficiency (7.5% to 0.3%) (Surun *et al.*, 2020). Prime editing is a promising genome editing strategy for common deletion mutations reported in LDs, including the c.1277_1278delAA in *GLA* and the c.3503_3504delTC in *GNPTAB*, the gene implicated in I-Cell

disease (Al-Jasmi *et al.*, 2013). Additionally, the Rec*Nci*I complex allele spans three mutations in a 50 bp region of *GBA1*. Anzalone *et al.* demonstrated prime editing with extended RT templates of up to 31 nucleotides (2019). Longer RT templates may result in a lower editing efficiency, but the possibility of utilizing an extended pegRNA to encompass all three mutations in the Rec*Nci*I complex allele should be explored. Reports of human pathogenic mutations edited using base and prime editing strategies are highly anticipated in the coming years.

1.3. Cell and gene therapies

1.3.1. Cell and gene therapies for LDs

Historically, cell therapies have encompassed the exogenous delivery of allogeneic or autologous cells to elicit a therapeutic effect (Biffi *et al.*, 2017). Therapeutic effect through cell therapy may be realized through either (1) repopulation of diseased or damaged cells, or (2) providing cells that excrete a soluble product to mitigate a disease state (Khaddour *et al.*, 2020; Deuse *et al.*, 2009). Initial examples of cell therapy include bone marrow transplants used to repopulate HSCs (Bach *et al.*, 1968). Stem cells are of particular interest in cell therapies due to their regenerative capacity and pluripotency, resulting in an ability to be differentiated into nearly any cell type (Kim and de Vellis, 2009). Pluripotent stem cells were historically obtained from human blastocysts in the form of human embryonic stem cells (ESCs) (Thomson *et al.*, 1998). Despite their immense clinical potential, ESCs are highly controversial due to their human embryonic origin. Ground-breaking research was conducted to determine the factors required to induce pluripotency in somatic cells, resulting in the discovery of four transcription factors, c-Myc, Klf4, Oct3/4, and Sox2 (MKOS), essential to ‘reprogramming’ cells to pluripotency

(Takahashi and Yamanaka, 2006). Resulting ‘reprogrammed’ cells were termed iPSCs, and displayed similar biochemical and gene expression profiles, and exhibited the trilineage differentiation abilities of ESCs (Takahashi and Yamanaka, 2006).

Since there are inherent risks associated with allogeneic HSCT, this form of cell therapy is reserved only for those with the most severe LDs and has been trialed primarily in MPSs (Armand *et al.,* 2014; Platt and Lachmann, 2009). Cell therapies are found to be more effective for some MPS disorders than ERT, and therapeutic benefit is more pronounced the earlier the cell therapy is administered (Shapiro *et al.*, 1995; Wang *et al.*, 2016; Peters *et al.*, 1998). In contrast to cell therapies, gene therapies introduce or modify genes within patient cells to bring about therapeutic effect (Kim and de Vellis, 2009). These two techniques intersect, however, for gene introduction or modification using genome editing strategies previously discussed can be applied *ex vivo* and then these cells can be delivered back to a patient for disease treatment (Biffi, 2016). A number of previously discussed preclinical and clinical methods for treating LDs, including the knock-in of *IDUA* and *IDS* at the albumin locus and of *GBA1* expression cassette knock-in at the CCR5 locus for treatment of MPS I, MPS II and GD, respectively, are fitting examples of emerging gene therapies (NCT02702115; NCT03041324; Scharenberg *et al.*, 2020). Current treatment options and preclinical/clinical genome editing strategies are listed in Table 1.1. For recent reviews on cell and gene therapies in LDs and in MPSs specifically, see Luciani *et al.* (2020) and Christensen *et al.* (2019).

Table 1.1 Current treatment options available for a number of lysosomal diseases (LDs). The potential use of genome editing strategies for the correction of pathogenic mutations or for the knock-in of a functional gene under a safe harbour locus is indicated for each LD listed. Modified from Christensen and Choy, 2017.

LD		Gene Affected	Current Treatment Options *	Example Drugs Available (Drug, *Company)*	Genome editing strategy under review (Strategy; Result)	References
Gaucher Disease (GD)		*GBA1*	ERT; SRT; PCT	Ceredase®, *Genzyme (Cambridge, MA, USA)*; Cerdelga®, *Genzyme (Cambridge, MA, USA)*; Mucoslovan®; *Boehringer Ingelheim (Biberach, Germany)*	CRISPR; Knock-in	Platt, 2014; Zimran *et al.*, 2013; Barton *et al.*, 1990; Weinreb *et al.*, 2002; Smid *et al.*, 2016; Scharenberg *et al.*, 2020
Sanfilippo Syndrome (MPS III)	A B C D	*SGSH* *NAGLU* *HGSNAT* *GNS*	SRT	Genistein ‡		De Ruijter *et al.*, 2012
Fabry		*GLA*	ERT	Fabrazyme®, *Genezyme (Cambridge, MA, USA)*	CRISPR; splice mutation deletion	Eng *et al.*, 2001; Germain *et al.*, 2015; Guce *et al.*, 2010; Chang *et al.*,2017
GM2-gangliosidoses (Tay Sachs or Sandhoff)		*HEXA; HEXB*	-	-	CRISPR; Knock in, CRISPR, HDR	Dersh *et al.*, 2016; Allende *et al.*, 2018; Ou *et al.*, 2020
I-cell disease		*GNPTAB*	-	-		Al-Jasmi *et al.*, 2013
Niemann-Pick C Disease (NPC)		*NPC1 or NPC2*	SRT	Zavesca®, *Actelion (Allschwil, Switzerland)*		Patterson *et al.*, 2007; Carstea *et al.*, 1997; Naureckiene *et al.*, 2000; Zech *et al.*, 2013
MPS I		*IDUA*	ERT	Aldurazyme®, *Genzyme (Cambridge, MA, USA)*	ZFN; Knock-in	Giugliani *et al.*, 2009; Ou *et al.*, 2019; NCT02702115
MPS II		*IDS*	ERT	Hunterase®, *CytoBioteck (Bogota, Colombia)*	ZFN; Knock-in	Sohn *et al.*, 2013; NCT03041324
MPS VI		*ARSB*	ERT	Naglazyme®, *Biomarin (San Rafael, CA, USA)*		Braunlin *et al.*, 2013
Pompe disease		*GAA*	ERT	Myozyme® *Genzyme (Cambridge, MA, USA)*		Lu *et al.*, 2003
Niemann-Pick A disease		*SMPD1*	-	-		Schuchman *et al.*, 2017; Schuchman *et al.*, 1992
Charcot-Marie-Tooth 1A		*PMP22*	-	-	CRISPR; Knock-out	Lee *et al.*, 2020
Spinal Muscular Atrophy		*SMN1*	-	-	CRISPR; HDR	Valetdinova *et al.*, 2019

* Only enzyme replacement therapy (ERT), substrate reduction therapy (SRT), pharmacological chaperone therapy (PCT) as current treatment options are indicated; ‡ genistein, a naturally-occurring isoflavone, has been shown to reduce urinary secretions of GAGs, but has yet to be tested at higher, clinically relevant doses for SRT in MPS III patients (De Ruijter *et al.*, 2012).

1.3.2. Intracerebral transplantation for GD and MPS IIIB

There is a precedent for intracerebral transplantation of allogeneic neural stem cells (NSCs) and neural progenitor cells (NPCs) to repopulating regions of the CNS, such as the subventricular zone, to give rise to new populations of neurons and glial cells to protect (Roitbak *et al.*, 2008) or treat ischemic injury in rodents (Zhao *et al.*, 2002). Furthermore, human glial grafts have been shown to attenuate disease symptoms in neonatal immunodeficient mice modelling Huntington's disease (HD) (Benraiss *et al.*, 2016). Intracerebral transplantation has been used in the clinic for both PD and HD; however, attenuation of disease symptoms requires long-term functionality of grafted cells, as well as immunosuppression to prevent GVHD (Benchoua and Onteniente, 2011). Additionally, two murine models exhibiting the LDs Niemann-Pick A disease and metachromatic leukodystrophy, show increased acid sphingomyelinase and arylsulfatase A activity, respectively, and reversal of lysosomal storage and cargo accumulation (Shihabuddin *et al.*, 2004; Meneghini *et al.*, 2017). Both models of intracerebral transplantation for LDs exhibit regional cross-correction (Shihabuddin *et al.*, 2004; Meneghini *et al.*, 2017). In conjunction with the aforementioned genome editing strategies, autologous patient-derived NSCs or NPCs edited for their pathogenic mutation may offer clinical benefit through intracerebral transplantation to patients with neurodegenerative GD or MPS IIIB.

1.4. Objective

At the outset of this project, the objective was to investigate genome editing strategies targeting pathogenic mutations in cells derived from patients with MPS IIIB and GD, since there is an unmet need in standard of care and therapeutic options for these patients currently (Sharpiro *et al.*, 2019; Bennett and Mohan, 2013). In the first phase of this project, we sought to identify mutations in a MPS IIIB cell line, previously reprogrammed to iPSCs in our lab (Turner, 2015), and characterize these cells for overall NAGLU protein expression. Additional investigations into droplet digital PCR (ddPCR) for GD mutation detection, with putative applications to non-invasive prenatal diagnosis (NIPD) were also explored. In the second phase, we sought to reprogram two skin FB cell lines, one each from a MPS IIIB and a GD patient, to pluripotency. At the outset of this experiment, the MPS IIIB cell line we chose had not yet been reprogrammed, thus of interest for future disease modeling of this genotype. In the third phase, we sought to edit mutations using CRISPR HDR in patient-derived iPSCs to demonstrate the ability to develop isogenic cell models for MPS IIIB and GD using this technology. As this is a rapidly progressing field, methods were adapted as more efficient genome editing strategies emerged. iPSCs were chosen as the cell type for genome editing due to their ability to form monoclonal colonies and ease of physical selection and screening in cell culture. However, iPSCs are difficult-to-transfect cells and genome editing efficiencies are reportedly low (Chatterjee *et al.*, 2011). Alternatively, genome editing in HEK293 cells and skin FBs was demonstrated. This method followed by reprogramming to iPSCs and monoclonal isolation may prove to be a more efficient method going forward.

2. Identification of a novel compound heterozygous mutation in MPS IIIB and screening applications for LD mutations

This chapter is adapted from the following publication:

Christensen, C.L., Ashmead, R.E., & Choy, F.Y.M. *Archives of Clinical and Biomedical Research,* **1**, 4 (2017).

Contributions: **CLC** designed the project, conducted experiments, and wrote the manuscript. RAE conducted experiments and edited the manuscript. FYMC supervised the project and edited the manuscript. All authors read the manuscript in full.

2.1. Introduction

Genotyping in MPS IIIB

Mutations in *NAGLU* are typically heterogenous, with private SNPs associated with single patients or families (Emre *et al.*, 2002; Schmidtchen *et al.*, 1998). However, p.Y140C, p.R297X and p.R626X missense and nonsense mutations have been reported in multiple families, constituting 25% of the total mutations found in one study (Beesley *et al.*, 1998). Another study reports p.R297X with an allelic frequency of 12.5% (Weber *et al.*, 1999). Furthermore, p.Y140C was reported in 18% of the tested alleles in heterozygosity with p.Y92H and p.L682R (Schmidtchen *et al.*, 1998). Knowledge of a patient genotype is important for predicting the overall disease severity and clinical course. For newborn screening and diagnosis, molecular genetic

analysis compliments enzyme activity and GAG accumulation assays to reduce false positives and predict phenotypic outcome (Donati *et al.*, 2018). Although next-generation sequencing (NGS) is the preferred method for diagnosing suspected multigene disorders, suspected single-gene disorders, such as GD and MPS IIIB, can still be rapidly and cheaply diagnosed at the molecular level through Sanger sequencing (Squeo *et al.*, 2020; Malaga *et al.*, 2019; Poswar *et al.*, 2019). Recently, advances in non-invasive prenatal testing (NIPT) directed towards genotyping pregnancies at risk for monogenic disease has made non-invasive early detection of rare disease possible through ddPCR (Camunas-Soler *et al.*, 2018). Knowledge of the possible fetal genotypes is required for this strategy, however (Camunas-Soler *et al.*, 2018).

Recently, elucidation of the NAGLU crystal structure has revealed that the majority of known pathogenic mutations alter amino acid residues involved in secondary structure elements (Birrane *et al.*, 2019). Of note, NAGLU residue Y140 was found to interact through hydrogen bonding with *N*-acetylglucosamine at the catalytic site (Birrane *et al.*, 2019). In homozygosity, p.R297X has been reported to result in very severe clinical course of MPS IIIB (Zhao *et al.*, 1998). Although p.Y140C has not yet been reported in homozygosity, when found in heterozygosity with a number of other mutations the clinical course has been reported as severe (Beesley *et al.*, 1998). Another private missense mutation in *NAGLU* that is referred to in this chapter is p.E153K, a mutation that was recently mapped to the interface of the beta-hairpin of domain I which stabilizes the tertiary structure of NAGLU (Schmidtchen *et al.*, 1998; Birrane *et al.*, 2019).

In addition to molecular genetic analysis, characterization of NAGLU enzyme activity and residual protein levels adds to the suite of analyses that compliment a clinical diagnosis. Nonsense

mutations may result in nonsense mediated mRNA decay or the truncated protein may be targeted for rapid degradation. Missense mutations resulting in misfolding may be targeted via the ubiquitin-proteasome pathway for degradation or may result in proper folding but lack enzyme activity. Enzyme structure and function outcomes can be predicted by a combination of immunoblotting and enzyme activity assays. Presently, *in silico* tools can be utilized to predict the impact of a mutation on enzyme structure and function (Punta and Ofran, 2008).

Restriction fragment length polymorphism (RFLP) analysis has been posited as a quick and relatively inexpensive method for genotyping genetic disease where potential pathogenic mutations have previously been elucidated (Soko *et al.*, 2018; Sharma and Changotra, 2018). Restriction sites are not always lost or gained in the presence of a pathogenic mutation, however. PCR primers can be designed that harbour a single nucleotide mismatch adjacent to a pathogenic mutation to create a restriction site in the resulting amplicon (Beutler *et al.*, 1990). Site-directed mutagenesis PCR strategies to create or remove restriction sites provides the ability to rapidly and simply genotype pathogenic mutations in *NAGLU* (Tseng *et al.*, 2002). Here, we report a novel compound heterozygous genotype, p.R297X/p.Y140C, with ~50% residual NAGLU protein as detected by immunoblotting and severely reduced enzyme activity. We also report two methods of site-directed mutagenesis to detect the p.R297X and p.E153K mutations in *NAGLU*.

Droplet digital PCR for GD

Amniocentesis and chorionic villus sampling (CVS) have been associated with a 0.86% and 0.74% increased risk of fetal loss, respectively (Kong *et al.*, 2006; Lau *et al.*, 2005). NIPT, in

addition to carrying no risk to the fetus, has been found to be more sensitive and have a lower false positive rate for detecting fetal aneuploidies (Norton *et al.*, 2015). Recently, progress has been made in developing NIPT for monogenic disorders using circulating cell-free fetal DNA (cffDNA) (Chang *et al.*, 2016; Chang *et al.*, 2018; Camunas-Soler *et al.*, 2018). ddPCR uses a partitioning technique to divide a single PCR reaction into thousands of nano reactions (Hindson *et al.*, 2011). Probes containing 3' fluorophores and 5' quenchers with sequence complementarity to a site of interest can then be used to detect the presence or absence of a rare allele upon hydrolysis of the probe (Liu *et al.*, 2006). Each reaction is measured from the presence or absence of fluorescence, resulting in the ability to generate a total quantitation of the overall proportion of DNA sequence containing the rare allele of interest (Vogelstein and Kinzler, 1999).

Recently, a technique for quantifying cffDNA in maternal blood plasma by analyzing SNPs with a minor allele fraction of >0.4 (Camunas-Soler *et al.*, 2018). ddPCR can then be used to accurately genotype the fetus based on the parent's carrier status (Camunas-Soler *et al.*, 2018). Using this technique, Camunas-Soler *et al.* accurately predicted the disease status of nine pregnancies (2018). Others have used variations of this technique to detect pregnancies with causative mutations for sickle cell disease (Barrett *et al.*, 2012), cystic fibrosis (Debrand *et al.*, 2015), and hemophilia (Tsui *et al.*, 2011). Genomic-based diagnosis of GD is complicated by the presence of a pseudogene, *GBAP1*, 16 kb downstream from *GBA1* that shares 96% sequence similarity with *GBA1* (Martinez-Arias *et al.*, 2001). Progress in NIPD for GD using cffDNA, NGS, and p.N370S-specific haplotyping has allowed for the accurate prediction of fetal genotype and avoids complications brought about by the *GBAP1* (Zeevi *et al.*, 2015). However, haplotyping-based tests run the risk of misdiagnosis due to meiotic recombination events (Altarescu *et al.*,

2008). Other *GBA1* genotyping methods include using functional gene-specific primers. Here, we describe a three-step PCR and ddPCR method for detecting p.N370S DNA spiked in a sample of wildtype DNA.

2.2. Materials and methods

2.2.1. General cell culture and gDNA extraction

MPS IIIB *NAGLU*$^{Y140C/R297C}$ skin FBs, cell line WG0421, was obtained from McGill-Montreal Children's Hospital Cell Repository. The cell line was submitted by Dr. Hayato Kihara from the Department of Pediatrics, UCLA. The clinical history of the patient is unavailable as Dr. Kihara died in 2007 and subsequently the patient's record was lost despite the department's efforts to retrieve this information. MPS IIIB *NAGLU*$^{E153K/E153K}$ skin FBs GM01426 was obtained from Coriell Biorepository. GD *GBA1*$^{N370S/N370S}$ skin FBs was a generous gift from Dr. Mia Horowitz, Tel Aviv University. FBs were cultured in Dulbecco's modified eagle medium (Gibco, 11960-044) supplemented with 1% (v/v) non-essential amino acids (Gibco, 11140-050) and 10% (v/v) fetal bovine serum (FBS; ThermoFisher, 16000-044) on T-75 flasks. Adherent FBs were disassociated from the T-75 flasks using 0.025% Trypsin-EDTA (Gibco, 154000-054).

2.2.2. PCR and genotyping

gDNA was extracted from skin FBs using the DNeasy® Blood & Tissue Kit (QIAGEN, 69504) as per manufacturer's protocol. *NAGLU* exons 1-6 were PCR-amplified using primers adapted from Schmidtchen *et al.* (1998). The thermal cycling conditions for exons 2 and 5 (595

bp and 434 bp amplicons harboring p.Y140C and p.R297X mutations, respectively) are as listed: initial denaturation at 94°C for 10 minutes, 37 cycles of denaturation at 94°C for 30 seconds, annealing at 55°C for 30 seconds, and elongation at 72°C for one minute, and a final elongation at 72°C for five minutes (Table 2.1). DNA amplicons were purified using the QIAquick PCR purification kit (QIAGEN, 28104) as per manufacturer's protocol. Sanger sequencing of these amplicons was performed by Eurofins Genomics.

Table 2.1. PCR primer sequences used to amplify *NAGLU*. Resulting amplicon sizes are indicated. Primer sets (1-6) are adapted from Schmidtchen *et al.* (1998). Table modified from Christensen *et al.*, 2017.

Exonic region		**Primer Sequence (5' – 3')**	**Amplicon size (bp)**
1	F	GGTCACACGCTCCCCACC	639
	R	ATTTGGGTGGCAGCGGCTCC	
2	F	CCCTGCCCATCTGTTAGACT	595
	R	GCACGTTGAAAGCACTTCTA	
3	F	AGCGCCCAGCACAAAGAAG	321
	R	CGGCCTAATACCACCCTTCCTG	
4	F	CGTGTATCCTGGGAGATGAGG	251
	R	GGCCCAGAGCTTAAGTTT	
5	F	AAACCAGGAGCTGTAGAGAAGT	434
	R	CTGCCTACCCCTACTGACATCT	
6-1	F	GGCCCTCTGTTTCATCACTC	1011
	R	TTGGCATAGTCCAGGATGTT	
6-2	F	CAACCGATCTGATGTGTTTG	904
	R	CAAGCGTGGCAGCAGTGACC	
MAF/MAR	F	ATCATCGGGAGCCTCTTCCT	142
	R	CCTCATAGACGGCAGTGGT	
2F/MM3R	F	AGAAGGGCCGAGTTTGGAG	208
	R	GCGCCATCCAGTCTGTCT	

2.2.3. Restriction fragment length polymorphism analysis

*Hpy*CH4III

Mutagenized primers were designed to create a cleavage site for the restriction enzyme *Hpy*CH4III (recognition site 5'-*A*CNGT-3'; normal sequence 5'-*C*CTGC; mutant sequence 5'-*C*CTGT-3'*) using forward primer (MAF) sequence 5'-ATC ATC GGG AGC CTC TT*A* CTG-3' and reverse primer sequence (MAR) 5'-CCT CAT AGA CGG CAG TGG T-3' for the amplification of gDNA (Table 2.1). The thermal cycling conditions using the MAF/MAR primer set are as listed: initial denaturation at 94°C for 10 minutes, 37 cycles of denaturation at 94°C for 30 seconds, annealing at 51°C for 20 seconds, and elongation at 72°C for one minute, and a final elongation at 72°C for five minutes. DNA amplicons were digested at 37°C for 60 minutes with 10 units of *Hpy*CH4III restriction enzyme (New England Biolabs, R0618S) and run on a 8% polyacrylamide gel, yielding 142 bp fragments for a normal human *NAGLU* sequence, and three fragments (142 bp, 122 bp and 20 bp) for a heterozygous p.Y140C/p.R297X *NAGLU* sequence. *Italicized sequence represents site of mutagenized base; underlined sequence represents the p.R297X mutation site.

*Bco*DI

Mutagenized primers were designed to create a cleavage site for the restriction enzyme *Bco*DI (recognition site 5'-*G*TCTCN-3'; normal sequence 5'-*A*TCTCN-3'; mutant sequence 5'-*A*TCTTN-3'*) using forward primer (2F) 5'-AGA AGG GCC GAG TTT GGA G-3' reverse

primer (MM3R) 5'-GCG CCA TCC AGT CT*G* TCT-3' for the amplification of genomic DNA (gDNA). The thermal cycling conditions using the 2F/MM3R primer set are as listed: initial denaturation at 94°C for 3 minutes, 37 cycles of denaturation at 94°C for 30 seconds, annealing at 53°C for 30 seconds, and elongation at 72°C for one minute, and a final elongation at 72°C for 5 minutes. DNA amplicons were digested at 37°C for 15 minutes with 5 units of *Bco*DI restriction enzyme (New England Biolabs, R0524S) then inactivated for 20 minutes at 65°C and run on a 8% polyacrylamide gel, yielding 175 and 35 bp fragments for a normal human *NAGLU* sequence, and a 208 bp undigested amplicon for a homozygous p.E153K *NAGLU* sequence. *Italicized sequence represents site of mutagenized base; underlined sequence represents the p.E153K mutation site.

2.2.4. ddPCR

gDNA was isolated from homozygous p.N370S patient skin FBs as previously described. A nested PCR strategy previously optimized in our lab was used to selectively amplify *GBA1* and not *GBAP1*. Functional gene-specific primers G10/G2II (Table 2.2) were used to amplify a 2.9 kb region of *GBA1*. LongAmp® *Taq* DNA Polymerase (New England Biolabs, M0323S) was used as per manufacturer's protocol. The thermal cycling conditions were as follows: initial denaturation at 94°C for 5 minutes, 34 cycles of denaturation at 94°C for 30 seconds, annealing at 52°C for 30 seconds, and elongation at 65°C for 3 minutes, and a final elongation at 68°C for 7 minutes. These 2.9 kb amplicons were used directly as template for a nested PCR with primers E8.5/S10 (Table 2.2) which was performed using Platinum Green Hot Start PCR Master Mix (Thermo Fisher Scientific, 13001013) with GC Enhancer as per manufacturer's protocol. The thermal cycling conditions were as follows: initial denaturation at 94°C for 3 minutes, 35 cycles

of denaturation at 94°C for 30 seconds, annealing at 53°C for 30 seconds, and elongation at 72°C for one minute, and a final elongation at 72°C for 5 minutes. All DNA oligonucleotides were obtained from Integrated DNA Technologies. Primers were used at a final concentration of 600 nM in all PCRs. Nested PCR resulted in 1.35 kb DNA amplicons, which were purified using the QIAquick PCR Purification Kit (QIAGEN, 28104) as per manufacturer's protocol. DNA concentration was confirmed via spectrophotometry and diluted to the digital range (1.0 – 2.0 x 10^4 DNA copies/reaction), assuming a molar mass/bp of 650 (g/mol)/bp and a fragment length of 1.35 kb results in 686 DNA copies/fg. A total mass of 22.5 fg was used for ddPCR experiments to achieve approximately 1.5 x 10^4 DNA copies/reaction. A probe with a FAM fluorophore was designed to target *GBA1* intron 8, and a probe with a HEX (VIC) fluorophore was designed to target the p.N370S mutation. Locked nucleic acids were introduced to the p.N370S probe complementary to bases c.1225-7.

ddPCR reactions were prepared with 22.5 fg total DNA mass, 1X QuantStudio™ 3D Digital PCR Master Mix v2 (Thermo Fisher Scientific, A26358), 250 nM REF probe (PrimeTime Eco Probe 5' 6-FAM/ZEN/3'IBFQ, Integrated DNA Technologies, Table 2.2), 250 nM N370S probe (Affinity Plus® Mini Probe 5' HEX™ 3'IB®FQ, Integrated DNA Technologies, Table 2.2), 900 nM each ddN370SF/ddN370SR primers (Table 2.2) to a final volume of 21.5 μL in dH_2O. QuantStudio™ 3D Digital PCR 20K Chip Kit v2 (Thermo Fisher Scientific, A26316) were loaded with 20 μL of reaction mixture using the QuantStudio™ 3D Digital PCR Chip Loader. Chips were thermal cycled on the ProFlex™ PCR System, 2 x flat, (Applied Biosystems™, LS4484078) with the following protocol: 96°C for 10 minutes, 39 cycles of 61°C for 2 minutes, 98°C for 30 seconds, then a final extension at 60°C for 2 minutes. ddPCR chips were imaged using the QuantStudio™

3D Digital PCR Instrument (Thermo Fisher Scientific, 4489084). Images were analyzed using QuantStudio™ Analysis Suite Software (Thermo Fisher Scientific). All samples were set to a quality score of 0.6, FAM threshold of 2,000 and VIC threshold of 2,500. This experiment was performed in triplicate.

Table 2.2 PCR primer and probe sequences for *GBA1* ddPCR. Locked nucleic acids are indicated by a "+" in front of the listed nucleotide.

Primer	Primer Sequence (5' – 3')	Amplicon Size
G10 G2II	GCC ATC TTC ACT CAC TGT AA CCT TTA ATG CCC AGG CTG AGC	2.9 kb
S10 E8.5	TTC AGC CCA CTT CCC AGA CC TGT GCA AGG TCC AGG ATC AG	1.35 kb
ddN370SF ddN370SR	TCT CCCACA TGT GAC CCT TAC CTA CA GGC AAG GTT CCA GTC GGT CCA	128 bp
REF probe	56-FAM/CCC CAG TGT /ZEN/TGC GCC TTT GTC TCT TTG C/3IABkFQ	N/A
N370S probe	5HEX/CCA CAT GGT ACA GGA G+G+C +TCT AGG G/3IABkFQ	

2.2.5. Western blot analysis

Human skin FBs were lysed using Mammalian Protein Extraction Reagent (M-PER®; Thermo Fisher Scientific, 78501). The total protein concentration of the cellular lysates was determined via a Bradford protein assay (Bradford, 1976). Total protein concentration of each cellular lysate was standardized prior to immunoblot analysis and 8.5 μg of total protein was loaded and separated on a 10% SDS-PAGE gel followed by an anti-NAGLU immunoblot for visualization of NAGLU. The apparent mass of NAGLU was calculated relative to a pre-stained protein standard (Bio-Rad, 1610374) which was run alongside the query protein samples. The membrane was marked where ladder bands denote protein size ranging 10-250 kDa. The primary antibody utilized in the immunoblot was a rabbit polyclonal antibody immunized against purified human NAGLU ab137685 (Abcam). Both the primary antibody and secondary goat anti-rabbit horseradish peroxidase conjugate antibody (Thermo Fisher Scientific) were applied at a 1:2,000 dilution. Relative band intensities were calculated using ImageJ v2 software (Rasband, 2008).

2.3. Results

2.3.1. Genotyping

PCR amplification of *NAGLU* followed by Sanger sequencing confirmed the p.Y140C/p.R297X genotype (Figure 2.1). DNA sequence analysis of exons 1, 3-4, and 6 showed normal human *NAGLU* sequence (data not shown). Genotyping was confirmed to be homozygous *NAGLU*$^{E153K/E153K}$ for GM01426 and homozygous *GBA1*$^{N370S/N370S}$ for GD skin FBs via Sanger sequencing (data not shown).

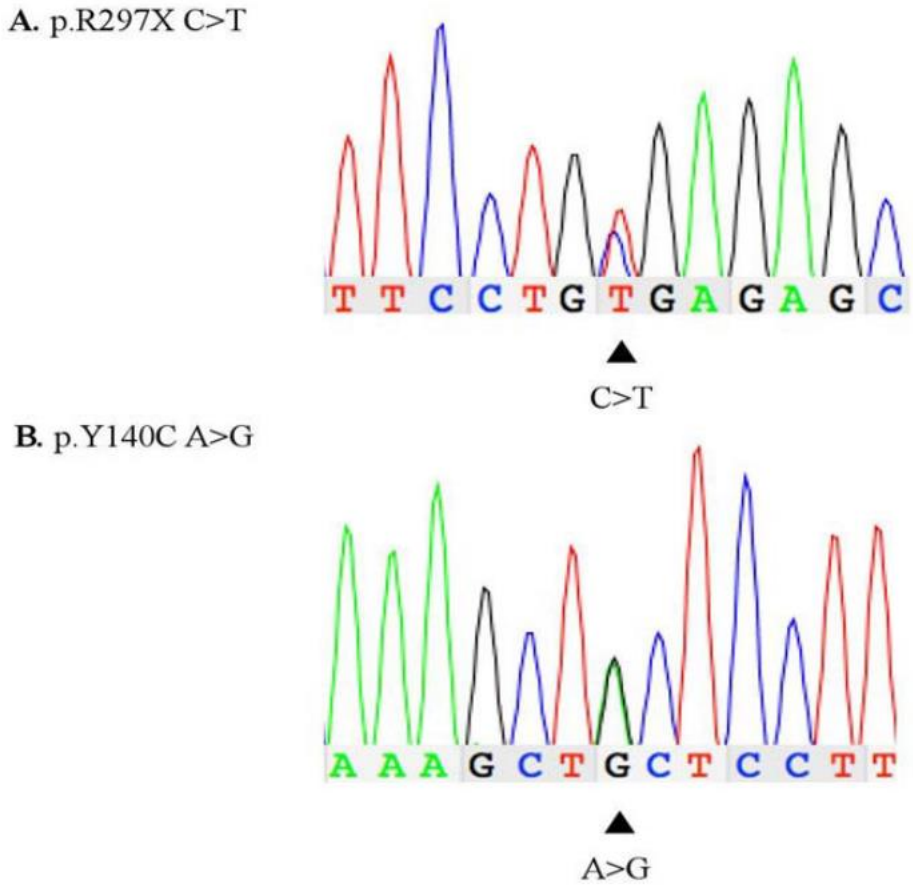


Figure 2.1. DNA chromatogram shows Sanger sequencing of exons 2 (B) and 5 (A) from patient WG0421. A. Heterozygous mutation p.R297X (C to T transition) at *NAGLU* cDNA nucleotide position 889. B. Heterozygous mutation p.Y140C (A to G transition) at *NAGLU* cDNA nucleotide position 419.

2.3.2. Restriction fragment length polymorphism analysis

p.R297X confirmation with *Hpy*CH4III

WG0421 gDNA amplified with a mutagenized primer and digested with the *Hpy*CH4III restriction enzyme resulted in the partial digestion of the resulting 142 bp amplicon to a 122 bp and 20 bp fragment, corresponding to the p.Y140C (undigested) and p.R297X (digested) alleles (Figure 2.2).

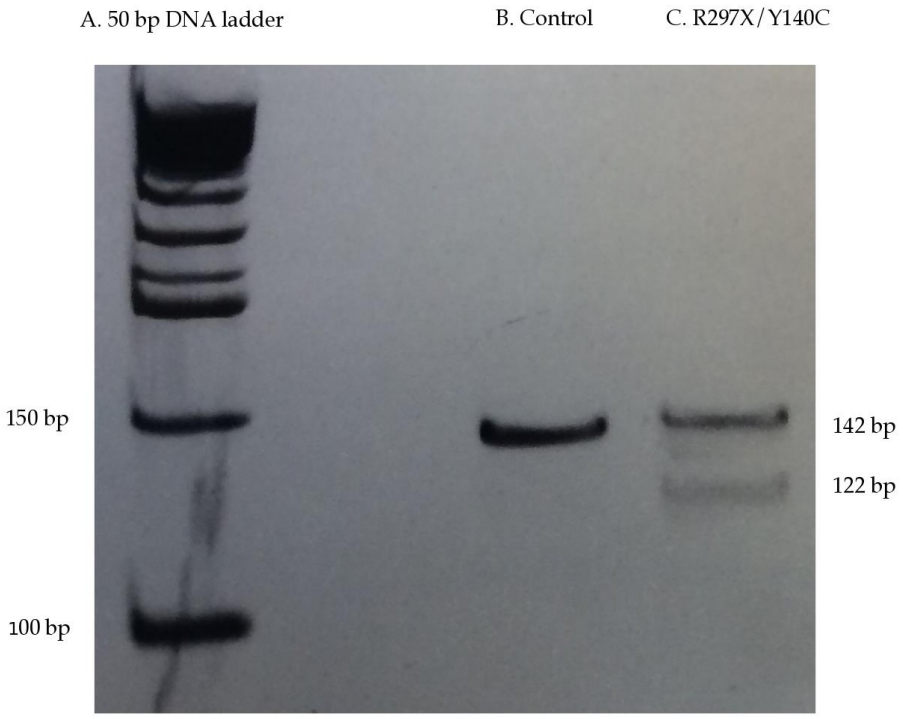


Figure 2.2. RFLP analysis for the genotyping of *NAGLU* mutation p.R297X in patient WG0421 on an 8% polyacrylamide gel electrophoresis. A: Marker, B: Control DNA amplicons, and C: WG0421 DNA amplicons. Amplicons (B and C) were generated using a mutagenized forward primer that creates a cleavage site for the enzyme *Hpy*CH4III when the p.R297X mutation is present, and a null site when the normal sequence is present. WG0421 has a compound heterozygous genotype (p.Y140C/p.R297X), thus creating a cleavage site for *Hpy*CH4III digestion

due to the p.R297X-mutated allele (122 bp and 20 bp fragments; 20 bp fragments not shown) and a normal p.R297 allele (142 bp band). Amplicons (1 µg DNA) were digested using *Hpy*CH4III restriction enzyme for 1 hour at 37°C, followed by a 65°C inactivation period.

p.E153K confirmation with *Bco*DI

Normal gDNA amplified with a mutagenized primer and digested with the *Bco*DI restriction enzyme resulted in full digestion of the 208 bp amplicon to 175 bp and 33 bp products, whereas GM01426 homozygous *NAGLU*$^{E153K/E153K}$ amplicons resulted in no digestion (Figure 2.3).

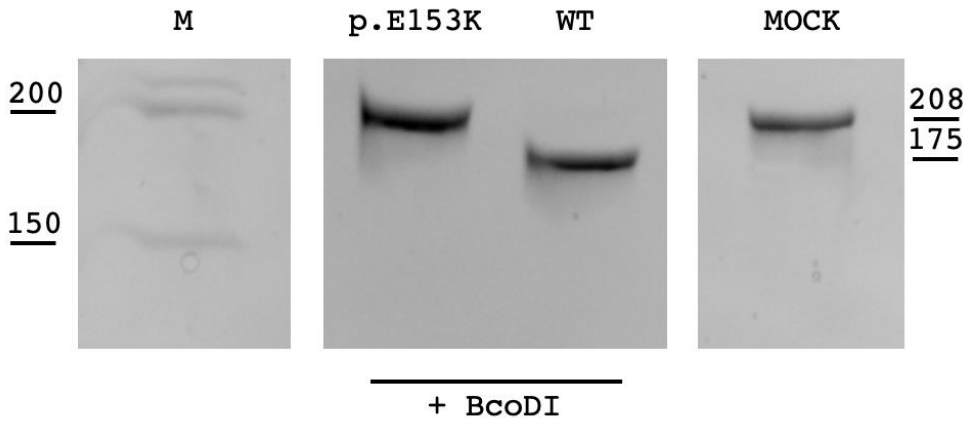


Figure 2.3. RFLP analysis for the genotyping of *NAGLU* mutation p.E153K in GM01426 on an 8% polyacrylamide gel electrophoresis. Amplicons were generated using a mutagenized reverse primer that creates a cleavage site for the enzyme *Bco*DI when the p.E153K mutation is absent. GM01426 has a homozygous genotype, thus abolishing the cleavage site for *Bco*DI digestion due to the p.E153K mutated alleles. This results in a 208 bp amplicon when the mutation is present, and a 175 bp and 33 bp (not shown) when the mutation is absent. Amplicons were digested using *Bco*DI restriction enzyme for 15 minutes at 37°C, followed by a 65°C inactivation period. Mock condition is a negative control for p.E153K alleles without addition of *Bco*DI.

2.3.3. Western blotting and enzyme activity assay

NAGLU-specific immunoblotting shows the presence of NAGLU at a size of approximately 70 kDa. ImageJ v2 (Rasband, 2008) densitometry indicates approximately 50% reduction in band intensity associated with the p.R297X/p.Y140C immunoblot vs. control (Figure 2.4).

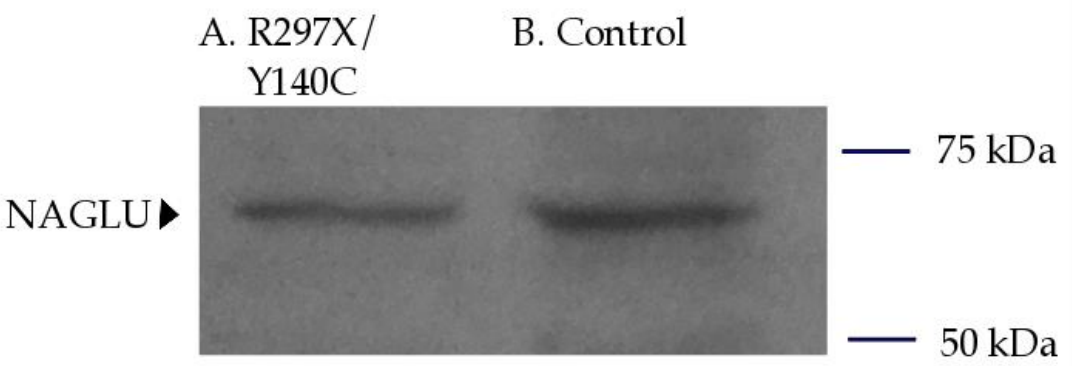


Figure 2.4. Total cellular protein harvested from confluent WG0421 (A) and control (B) FBs and separated on a 10% SDS-PAGE gel followed by an anti-NAGLU immunoblot for visualization of NAGLU. A total of 8.5 μg of protein were applied per lane. Both primary and secondary antibodies were applied at a 1:2,000 dilution. Lane A: total protein lysate from patient WG0421 FBs (p.Y140C/p.R297X genotype). Lane B: total protein lysate from control FBs. NAGLU-specific banding is shown in both lanes, corresponding to an apparent 70 kDa product. Relative protein banding intensities (A vs. B) were quantified and compared using ImageJ v2 software (Rasband, 2008) by measuring the area under the curve (Hansen *et al.*, 2012). A pre-stained protein standard (Bio-Rad) was included with the SDS-PAGE and used to mark the membrane, where ladder bands denote protein size ranging 10-250 kDa (not shown) and where the apparent NAGLU protein band fell between 50 and 75 kDa.

NAGLU activity assay was performed at Centre Hospitalier Universitaire Sainte-Justine in Montreal, QC. WG0421 patient FBs yielded 0.1 nmoles/mg/hr of NAGLU enzyme activity in

comparison to 16 nmoles/mg/hr NAGLU activity in normal FBs, corresponding to about 0.6% of normal (data not shown).

2.3.4. ddPCR for p.N370S detection

ddPCR using a reference probe containing a FAM fluorophore and a p.N370S-specific probe containing a HEX (VIC) fluorophore resulting in FAM-positive event detection, with an error of 1.16 ± 0.23% associated with FAM/VIC-positive event detection when gated for a VIC threshold of 2,500 and queried with only normal human DNA amplicons. FAM/VIC-positive events were detected in 99.70 ± 0.02% of the total FAM-positive events when 100% homozygous p.N370S DNA amplicons were queried. Spiking of 30%, 10%, 5%, and 0.1% p.N370S DNA amplicons in a solution of normal human DNA amplicons resulted in 29.03 ± 1.08%, 10.86 ± 0.40%, 5.96 ± 0.04%, and 6.28 ± 2.09%, respectively (Figure 2.5; Supplementary Table 2.1). Experimental conditions were performed in triplicate. Example event plots for each representative condition are shown in Figure 2.6.

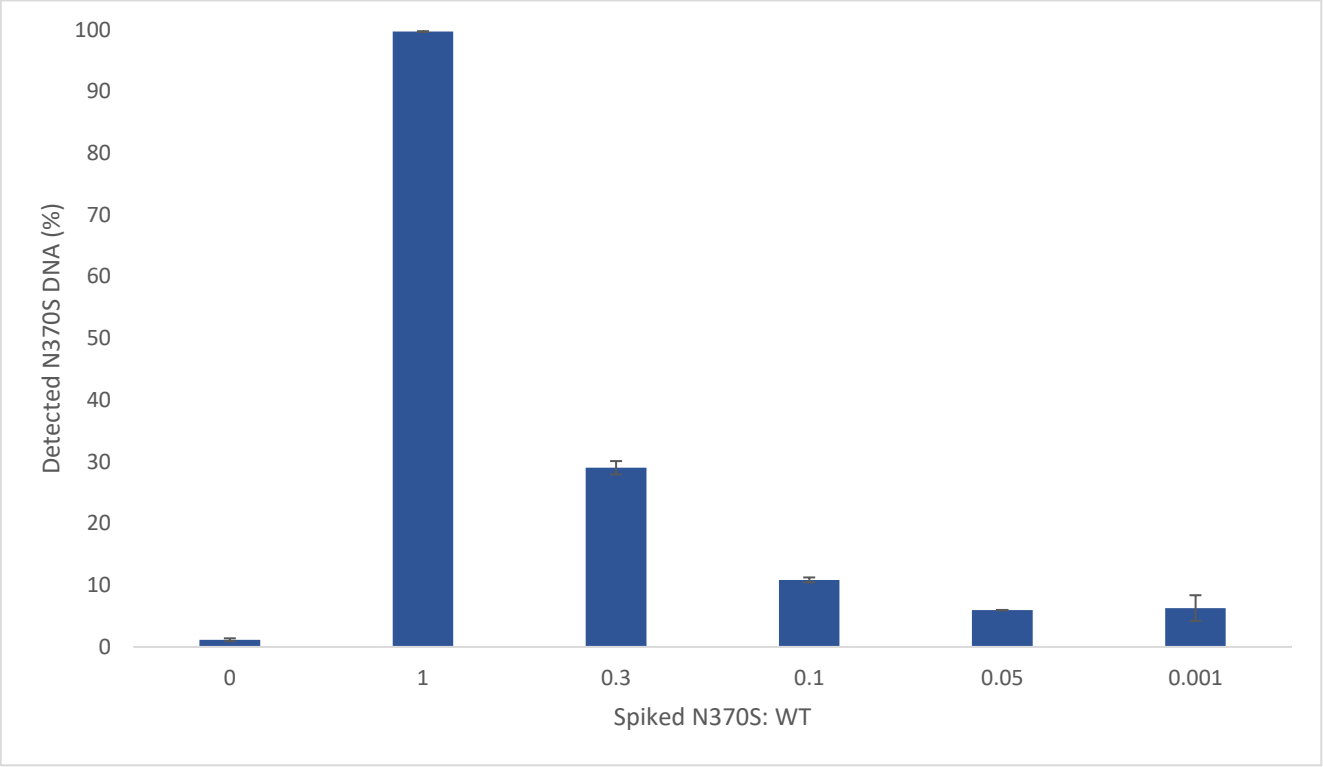


Figure 2.5. Graph depicting p.N370S FAM/VIC positive events out of total FAM-positive events detected (%) versus the spiked ratio of N370S: wildtype (WT) DNA. Only WT (0% spiked p.N370S), 1.16 ± 0.23%; 100% spiked p.N370S, 99.70 ± 0.02%; 30% spiked p.N370S, 29.03 ± 1.08%; 10% spiked p.N370S, 10.86 ± 0.40%; 5% spiked p.N370S, 5.96 ± 0.04%; 0.1% spiked p.N370S, 6.28 ± 2.09%. Experiment was performed in triplicate.

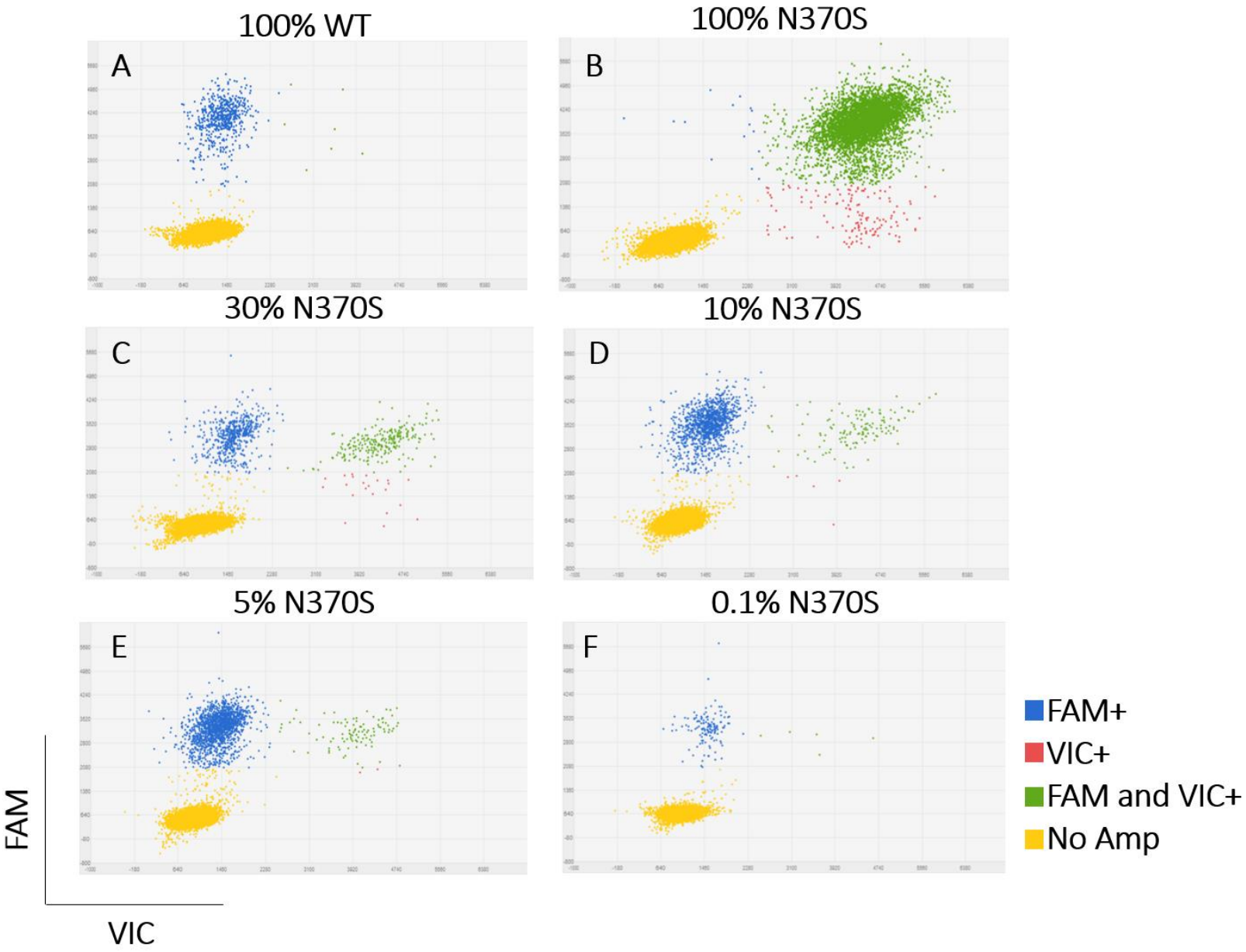


Figure 2.6. ddPCR 2D plots show reference (FAM) and p.N370S (VIC) probe binding in various p.N370S homozygous template-spiked conditions. Non-specific reference probe binding leads to FAM fluorophore release (single-positive, blue events), mutation-specific probe binding leads to VIC fluorophore release (double-positive, green events), and no probe binding is evidenced by a lack of fluorophore release (double negative, yellow events). 2D plots were generated using QuantStudio™ Analysis Suite Software (Thermo Fisher Scientific) with a quality score of 0.6, FAM threshold of 2,000, VIC threshold of 2,500. p.N370S DNA was present at A. 0% B. 100% C. 30% D. 10% E. 5% and F. 0.1% of total DNA sample resulting in 1.16 ± 0.23%,

99.70 ± 0.02%, 29.03 ± 1.08%, 10.86 ± 0.40%, 5.96 ± 0.04%, and 6.28 ± 2.09% positive events, respectively. Experiment was performed in triplicate.

2.4. Discussion

Genotypes causative for MPS IIIB are often heterogenous and specific only to single patients or families (Emre *et al.*, 2002; Schmidtchen *et al.*, 1998). It is important to have a molecular genetic diagnosis to predict disease severity and clinical course. Here, we sought to genotype and determine the residual NAGLU protein levels in patient skin FBs with a very low (0.6% of normal) NAGLU enzyme activity. Through PCR amplification of the *NAGLU* gene and subsequent Sanger sequencing, we determined the patient genotype to be a novel compound heterozygous mutational complement, p.Y140C and p.R297X, two relatively common mutations located in exon 2 and exon 5, respectively (Beesley *et al.*, 1998). In the current study, immunoblotting indicated approximately 50% of normal NAGLU protein levels, compared to normal control skin FBs; however, given the lack of a control loading protein these results should be considered qualitative rather than quantitative. The low residual NAGLU activity suggests that p.R297X and p.Y140C both contribute to a severe form of MPS IIIB, and we propose that the allele harbouring the p.R297X nonsense mutation results in either an unstable messenger RNA (mRNA) transcript, or the production of a truncated protein that is rapidly degraded. Future studies should focus on quantitatively determining the impact of the p.Y140C mutation on overall residual NAGLU protein levels and elucidating NAGLU enzyme activity in homozygous p.Y140C skin fibroblasts.

The allele harbouring p.Y140C is proposed to produce a catalytically inactive NAGLU protein, thus resulting in approximately 50% of normal NAGLU protein levels. Comparatively, a study by Yogalingam *et al.* aimed to characterize the compound heterozygous mutation p.R297X/pF48L (nonsense/missense mutation) through *NAGLU* cDNA expression in Chinese hamster ovary (CHO) cells to determine the role of both mutations in producing the patient's observed attenuated phenotype (2000). p.R297X was shown to result in undetectable NAGLU activity and a truncated 34 kDa NAGLU polypeptide, which was rapidly degraded (Yogalingam *et al.*, 2000). Expression of the p.F48L mutation produced a 80 kDa precursor with 3.75% residual NAGLU activity, which was eventually degraded as well (Yogalingam *et al.*, 2000). In the current study, NAGLU immunoblot analysis indicates the apparent size of NAGLU to be 70 kDa, which falls within range of previously reported NAGLU size (67-83 kDa), where this range is likely due to tissue-specific protein modifications (Jantzen *et al.*, 2013; Zhao *et al.*, 2000; Sasaki *et al.*, 1991; Weber *et al.*, 2001).

In order to confirm the role of the p.Y140C/p.R297X mutations, expression of both the p.R297X and p.Y140C alleles separately from each other is necessary. Although phenotypic data is not available for patient WG0421, it is speculated that this patient presented with a severe, unattenuated, form of MPS IIIB due to the near zero (0.1 nmol/mg/hr) residual NAGLU activity detected. In a previous study by Piotrowska *et al.* on the correlation between severity of MPSs (Hunter and Sanfilippo A and B syndromes) and the residual enzyme activity and efficiency of GAG synthesis, it was postulated that a severe MPS IIIB phenotype is observed in patients lacking residual NAGLU activity, regardless of GAG synthesis efficiency (2009). Based on the study by Piotrowska *et al.*, the near zero residual enzyme activity measured in the current study is likely to

result in a severe phenotype in the patient (2009). In a recent study to discern attenuated and severe MPS IIIB phenotypes based on biochemical assays in FBs, Meijer *et al.* determined an optimal cut-off of 0.58 nmol/mg/hr, thus placing our patient (at 0.1 nmol/mg/h) below the cut-off for an attenuated form of the disease (2016). More recently, this group interestingly reported a higher NAGLU activity when patient skin FBs are cultured at 30°C instead of 37°C (Meijer *et al.*, 2017). They also show through western blot analysis that homozygous p.R297X skin FBs exhibit no detectable mature NAGLU protein whatsoever (Meijer *et al.*, 2017). Additionally, p.R297X *NAGLU* mRNA levels are greatly reduced in comparison to missense mutations p.S612G, p.R643C, and compound heterozygous p.R565Q/p.A72_G79dup8. However, they curiously chose qPCR primers that target *NAGLU* at cDNA position 960, 71 bases downstream of the c.889 C>T (p.R297X) mutation in exon 5 (Meijer *et al.*, 2017). This may confound these results by excluding the presence of a truncated *NAGLU* mRNA transcript (Meijer *et al.*, 2017). If accurate, this is further evidence that the p.R297X mutation results in either an unstable mRNA transcript or a truncated protein that is rapidly degraded (Meijer *et al.*, 2017).

Since there are no existing inherent restriction sites to allow for a diagnostic restriction endonuclease digest for p.R297X (c.889 C>T), we have employed site-directed mutagenesis by using mismatch PCR to create a restriction site for the *Hpy*CH4III restriction enzyme, where a restriction site is present when the p.R297X mutation is present. This diagnostic restriction digest can be used for confirmation of the p.R297X mutation where this mutation is suspected. An inherent restriction site does not exist for the p.Y140C, although a similar site-directed mutagenesis may create a restriction site for diagnostic restriction digest. A novel diagnostic restriction endonuclease digest for p.E153K (c.457 G>A) was also developed, wherein site-

directed mutagenesis by using mismatch PCR was employed to create a restriction site for *Bco*DI when the normal c.457 G nucleotide is present, and abolish this site when the c.457 A mutation is present. Although the p.E153K mutation has only been reported once (Schmidtchen *et al.*, 1998), this mutation has recently been mapped to the beta-hairpin of domain I and postulated to impact the tertiary structure of the NAGLU protein (Birrane *et al.*, 2019). Other mutations mapped to the beta-hairpin of domain I may also lead to decreased protein stability. Recently, an insertion mutation, c.455_456insA (E153Rfs*39), was reported as causing a frame shift resulting in the glutamic acid residue at position 153 to be changed to arginine and leading to a stop codon downstream (Mohammed and Fateen, 2019). The site-directed mutagenesis followed by *Bco*DI restriction digest to identify the p.E153K mutation described here is predicted to also recognize this insertion mutation, resulting in the abolishment of the restriction site. Therefore, this method may be used as a relatively quick and inexpensive way to confirm the molecular genotype of patients suspected of having either the p.E153K or p.E153Rfs*39 mutation.

In this chapter, a ddPCR design that lays the foundation for a NIPD technique to detect the most common mutation in *GBA1*, p.N370S, has also been described. A reference and rare allele (p.N370S-targeting) probe were designed to target intron 8 and exon 9 of *GBA1*, respectively. Locked nucleic acids were introduced in the rare allele probe oligonucleotide sequence complementary to the c.1225-7, thus increasing the stability and selectivity of Watson-Crick base pairing at that location (Singh *et al.*, 1998). Typically, ddPCR is performed on gDNA samples utilizing primers that flank a short 70-150 bp sequence of interest that encompasses the probe binding sites. However, in the case of *GBA1*, a pseudogene with 96% sequence similarity complicates methods utilizing gDNA and necessitates a cursory PCR amplification step (Martinez-

Arias *et al.*, 2001). To achieve selective *GBA1* amplification, we used a reverse primer (G2II) that binds to a 55 bp sequence present downstream of exon 9 that is deleted in *GBAP1*. A nested PCR is then performed to create a 1.35 kb amplicon flanking c.1226. Nested c.1226 A>G-containing amplicons diluted to within the digital range were used to spike normal nested DNA amplicons, generating a range of 'rare' mutant alleles, from 0.1 – 100%. The rare allele probe is highly sensitive to c.1226 A>G-containing DNA (99.70 ± 0.02%), but not to normal DNA (1.16 ± 0.23%).

The accuracy of detection decreased as the percent spiked rare allele within the total sample decreased, with an apparent event detection threshold of above 5% (Figure 2.5 –.6). However, given that the number of total FAM and/or VIC positive events are approximately less than one third of the total loaded DNA, this indicates that further optimization of PCR parameters, such as annealing temperature and number of cycles, may increase the sensitivity and limit of detection of this assay. Additionally, in this study the variability of the number of total events differs greatly across conditions. Specifically, the 0.1% spiked p.N370S condition exhibits a substantial false positive rate given the overall detected FAM/VIC positive event percentage of 6.28 ± 2.09% (Figure 2.6). The FAM/VIC positive event percentages calculated here are represented as a proportion of total positive events. Therefore, the FAM/VIC positive event percentage may be overexaggerated when the number of total events is limited in any given condition. Across our 0.1% spiked p.N370S trials, we had a surprisingly low level of total events (88 ± 24) in comparison to our negative control condition (549 ± 19), exhibiting an approximately 6x reduction in total events. However, we found a rather constant number of FAM/VIC positive hits (5 ± 2 vs. 7 ± 2) between the 0.1% spiked p.N370S and negative control conditions, respectively, indicating that we may be able to reduce our level of detection by increasing specificity of our reference probe.

The total amount of cffDNA averages 13% of the total cell-free DNA in maternal plasma, but this varies depending on a number of factors, including gestational age (Hu *et al.*, 2019). Fetal fraction in gestational ages 11-13 weeks can be less than 4%, and this serves as the threshold for most non-invasive prenatal screening methodologies (Ashoor *et al.*, 2013). Given this threshold, our ddPCR method will need to be further optimized to increase accuracy of detection below 10% of rare allele fraction. Recent studies have developed techniques that increase the amount of total cffDNA present in the sample, which may also mitigate issues of reduced sensitivity at lower concentrations of total cffDNA (Hu *et al.*, 2019; Chang *et al.*, 2018). ddPCR offers an alternative option to amniocentesis and CVS, with no associated risk of fetal loss (Kong *et al.*, 2006; Lau *et al.*, 2005). Going forward, this ddPCR technique will be tested using samples of maternal plasma enriched for cffDNA. This strategy of detecting c.1226 A>G spiked DNA in a pool of normal DNA through ddPCR is foundational to developing future detection techniques to identify this common pathogenic mutation in GD. Going forward, development of NIPT strategies should focus on common mutations in monogenic disorders, such as p.L444P and p.N370S in *GBA1*, for this will provide putative clinical benefit to the largest number of patients, especially within founder populations such as Ashkenazi Jewish (Kolodny *et al.*, 1982).

2.5. Conclusions

Molecular diagnostic techniques are essential for identifying patient genotype and inferring predicted disease severity and clinical course. In addition to sequencing techniques, site-directed mutagenesis followed by RFLP offers a relatively inexpensive and rapid method to genotype a patient when (1) a disease-causing mutation is prevalent or (2) the patient genotype can be

predicted from a close relative's genotype. Here, we report a novel compound heterozygous genotype shown to result in ~50% of normal NAGLU protein and limited NAGLU enzyme activity (0.6%). Additionally, we demonstrate two methods of site-directed mutagenesis followed by RFLP for detecting MPS IIIB mutations p.E153K and p.R297X. Lastly, we demonstrate a ddPCR strategy for detecting the p.N370S mutation in *GBA1* using a rare allele probe with locked nucleic acids.

2.6. Supplementary information

Supplementary Table 2.1. Average FAM+VIC positive events in a range of p.N370S DNA amplicon spiked conditions. Data was analyzed with QuantStudio™ Analysis Suite Software (Thermo Fisher Scientific), with a quality score of 0.6, FAM threshold of 2,000 and VIC threshold of 2500. *Outliers have been excluded from average percent positive event calculation.

Template	FAM	FAM + VIC	Positive events (%)	Average positive events ± SE (%)
100% WT	390	36	8.45*	1.16 ± 0.23
	530	5	0.93	
	568	8	1.39	
100% N370S	11	3699	99.70	99.70 ± 0.015
	15	5403	99.72	
30% N370S	10	3027	99.67	29.03 ± 1.08
	550	237	30.11	
	647	251	27.95	
10% N370S	1125	237	17.40*	10.86 ± 0.40
	1184	134	10.17	
	1321	161	10.86	
5% N370S	1057	138	11.55	5.96 ± 0.035
	1238	79	5.99	
	1622	102	5.92	
0.01% N370S	123	76	38.19*	6.28 ± 2.09
	69	8	10.39	
	136	5	3.55	
	58	3	4.92	

3. Reprogramming MPS IIIB and GD patient fibroblasts to pluripotency

3.1. Introduction

iPSCs are of interest to the regenerative medicine field due to their applications in gene therapies, disease modeling, and drug screening. Stem cells in general are of particular interest due to their innate renewal capacity and ability to be differentiated into virtually all tissue types (Takahashi and Yamanaka, 2006). This property of stem cell renewal capacity was first demonstrated in 1960 by McCulloch and Till with their seminal findings of hematopoietic cell replenishment after radiation in mice, outlining the discovery of HSCs (1960). Since this discovery, HSCs have been used as a method to treat non-malignant hematologic disorders (Mahmoud *et al.*, 2015), autoimmune disease (Tyndall *et al.*, 1997), solid tumours (Gratwohl *et al.*, 2004), and metabolic disorders including a number of LDs (Biffi, 2017). It was once debated whether non-hematologic disorders would benefit from HSCT; however, a mechanism through which acid hydrolases can transverse the plasma membrane has allowed for systemic therapeutic effect in some LDs (Aldenhoven *et al.*, 2017; Fratantoni *et al.*, 1968).

iPSCs circumnavigate issues associated with allogeneic transplantation since they are derived non-invasively from the patient themselves. The 2006 discovery of the four transcription factors, c-Myc, Klf4, Oct3/4, Sox2 (MKOS) that allow for reprogramming of somatic cells to pluripotency has allowed for the establishment of patient-derived iPSCs *in vitro* to become routine (Takahashi and Yamanaka, 2006). When iPSCs are generated from a patient's own somatic cells, patient-specific disease modelling, drug testing, and the generation of cells for transplantation without immune rejection risk is possible (see chapter 1 for discussion; Nadig, 2009). In the past,

the study of neurodegenerative disease required animal models. With iPSC technology, neurodegenerative pathways can be studied in human iPSC-derived neuronal cell cultures, which can provide further insight into the disease mechanisms and pathways. The recent discovery of the genome editing technique, CRISPR-Cas9, has provided further disease modeling possibilities through editing mutations in patient cell lines (Jinek *et al.*, 2012). Although the efficacy, safety, and differentiation potential of iPS-derived NPC grafts are still being elucidated (Benchoua and Onteniente, 2011), early studies have shown that intracerebral transplantation of NPCs or NSCs mitigates neuropathology due to lysosomal storage defects in mice (Shihabuddin *et al.*, 2004; Fukuhara *et al.*, 2006).

Methodologies for generating iPSCs from FBs include viral integrating methods such as lentiviral and retroviral transduction (Yu *et al.*, 2007; Takahashi and Yamanaka, 2006), as well as non-integrating methods like episomal plasmid vectors and Sendai viral (SeV) techniques (Okita *et al.*, 2011; Fusaki *et al.*, 2009). Although viral integrating methods are more efficient at generating iPS colonies *in vitro*, non-integrating methods lack risks associated with viral methods, such as transgene integration in the genome, a concern for use in clinical applications. Since SeV is an RNA virus and does not enter the nucleus, it is diluted from the cytoplasm typically within 10 passages (MacArthur *et al.*, 2012; Ban *et al.*, 2011). iPSCs are typically characterized for their trilineage differentiation to embryonic germ layers (ecto-, meso-, and endoderm) and expression of pluripotency markers. Alkaline phosphatase (ALP), a membrane-bound enzyme responsible for the hydrolysis of phosphate under alkaline conditions, is highly expressed in iPSCs (Stefkova *et al.*, 2015; Singh *et al.*, 2012). ALP live staining (APLS) is commonly used as a first line of screening for pluripotency in newly emerging iPS colonies (Singh *et al.*, 2012). A number of pluripotency markers have been used in the literature to assess pluripotency, including Nanog,

Oct4, stage-specific embryonic antigen (SSEA)-3, SSEA-4, TRA-1-60, and TRA-1-81 but are negative for other cell-surface markers such as SSEA-1 (Kim *et al.*, 2009; Boulting *et al.*, 2011; Ojima *et al.*, 2015). Generating teratomas under the skin of immunocompromised mice by injecting iPSCs is regarded as the "gold standard" for confirming pluripotency (Gutierrez-Aranda *et al.*, 2010).

The accumulation of GAGs in the MPSs are well known to result in cellular disease pathology due to increased lysosomal compartment size (Wilkinson *et al.*, 2012; de Araujo *et al.*, 2020) and increasing lysosome number impacting normal cellular functioning (Gadola *et al.*, 2006). GAGs have a number of biologic activities, including key interactions with cytokines and growth factors. For example, heparin and chondroitin sulfate (Balasubramanian and Ramanathan, 2000), as well as HS and hyaluronic acid (Fernandez-Botran *et al.*, 1999) have all been shown to cause a structural change in interferon γ, which may suggest a function of presenting this cytokine to its receptors (Fernandez-Botran *et al.*, 1999). Additionally, heparin and HS have been shown to decrease the signal transducer and activator of transcription (STAT) signalling pathway through inhibition of interleukin 27 (Cavé *et al.*, 2020). Therefore, it is also important to explore the impact of GAG accumulation in biologic functions including cell signaling pathways.

HS, the mucopolysaccharide that accumulates in MPS IIIB, is found as bound HSPG chains at the cell surface and play a critical role in signaling pathways involved in development (Lin, 2004), tumourigenesis (Sasisekharan *et al.*, 2002) and in maintenance of pluripotency and stem cell fate (Kraushaar *et al.*, 2013). HS interacts with fibroblast growth factor (FGF)-2 at the cell surface, not only acting as a reservoir to protect this mitogen from degradation, but also reducing its bioavailability (Lemonnier *et al.*, 2011; Quarto and Amalric, 1994). Canonical FGF cell signalling families require HS as a cofactor to bind and activate FGF receptors, forming a ternary

FGF-FGFR-HS complex that initiates a phosphorylation of tyrosine residues (Ornitz and Itoh, 2015). FGF2 signaling plays a role in maintaining stem cell pluripotency through mitogen-activated protein kinase (MAPK), and phosphatidylinositol 3 kinase (PI 3-K)/Akt (protein kinase B) pathways (Mossahebi-Mohammadi *et al.*, 2020).

In addition to abundant HS sequestering FGF2, partially degraded or improperly sulfated HSPGs may also negatively impact cell proliferation through decreased ability to interact as a cofactor with FGF2 (Pan *et al.*, 2005; Lemonnier *et al.*, 2011). Abnormal 6-O-sulfation of HSPGs in multipotent progenitor cells (MPCs) from patients with MPS IH resulted in reduced proliferation and survival of these cells (Pan *et al.*, 2005). They found that the addition of HS to the culture media rescued FGF2-induced proliferation and survival of patient MPCs (Pan *et al.*, 2005). Additionally, partially degraded HS has been postulated to impair proliferation of emerging MPS IIIB iPSCs where low reprogramming efficiencies have been observed (Lemonnier *et al.*, 2011). We have also observed impaired proliferation of MPS IIIB iPSCs, manifesting with increased cell death and spontaneous differentiation to spheroids resembling embryoid bodies (EBs). Curiously, EBs, which consist of endo-, meso-, and ectoderm, are routinely generated as a means to assess the pluripotency of iPSCs by increasing the 2-mercaptoethanol concentration and removing FGF2 (Son *et al.*, 2011). We postulate that supplementing additional FGF2 into the culture media may mitigate cell death and spontaneous differentiation arising from reduced bioavailability or increased degradation of FGF2 (Figure 3.1).

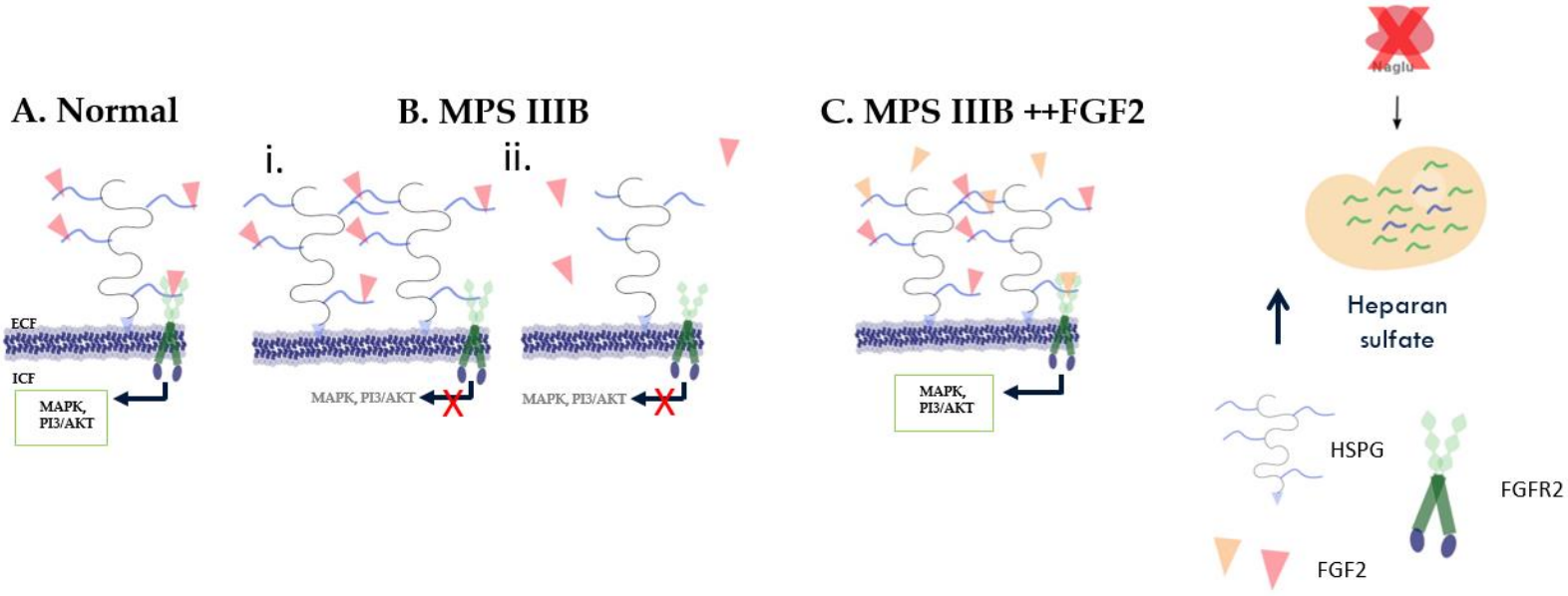


Figure 3.1 Schematic of HS and FGF2 interactions at the plasma membrane under normal (A) and MPS IIIB (B) cellular conditions. Under normal cellular conditions, heparan sulfate proteoglycans (HSPGs) present in extracellular matrices bind free FGF2 and are required cofactors for FGF2 docking. Thus, disorders of heparan sulfate (HS) regulation and processing are postulated to impact FGFR2 and -3 phosphorylation cascades and their downstream targets, including the MAPK/ERK pathway, a critical pathway for promoting reprogramming to and maintenance of induced pluripotent stem cells (iPSCs). When HS accumulates in MPS IIIB, we postulate that higher FGF2 concentration is required to saturate free HS and bound HSPGs (**C.**) **B.** (**i.**) represents the hypothesis that increased HSPG accumulation leads to sequestration of FGF2, and (**ii.**) represents the hypothesis that improperly degraded HS leads to an increase in FGF2 degradation.

Here, we describe the reprogramming of *NAGLU*$^{E153K/E153K}$ and *GBA1*$^{N370S/N370S}$ human skin FBs to iPSCs. Inherent disease pathology is a critical consideration when utilizing iPSCs as an intermediary cell type for disease modelling or therapeutics. Based on observations of increased cell death and spontaneous differentiation in comparison to GD and normal iPSCs, we also sought

to explore the impact of supplemental FGF2 on overall cell confluency for early passage MPS IIIB iPSCs.

3.2. Materials and methods

3.2.1. Cell culture and reprogramming

Fibroblast culture

MPS IIIB *NAGLU*$^{E153K/E153K}$ skin FBs GM01426 was obtained from Coriell Biorepository. GD *GBA1*$^{N370S/N370S}$ skin FBs was a generous gift from Dr. Mia Horowitz, Tel Aviv University. The cells were cultured in Dulbecco's Modified Eagle Medium (D-MEM; Gibco, 11960-044) supplemented with 10% (v/v) FBS (ThermoFisher, 16000-044) and 1% (v/v) non-essential amino acids (Gibco, 11140-050) at 37°C/5% CO_2. Media was replenished every 3-4 days and cells were passaged or frozen at 75-85% confluence. FBs were disassociated using 0.025% Trypsin-ethylenediaminetetraacetic acid (Gibco, 154000-054). HEK293 cells were disassociated mechanically. Freezing media consisted of culture media with 10% (v/v) dimethyl sulfoxide (Sigma Aldrich, D8418).

Skin fibroblast reprogramming

NAGLU$^{E153K/E153K}$ and *GBA1*$^{N370S/N370S}$ skin FBs were reprogrammed after the ninth and sixth passage, respectively. Skin FBs were grown to 75% confluence. Skin FBs are transduced using the CytoTune™-iPS 2.0 Sendai Reprogramming Kit (Thermo Fisher Scientific, A16517) with three reprogramming vectors containing hc-Myc, hKlf4, and a combination of hKlf4, hOct4,

and hSox2 (KOS) as per manufacturer's protocol. Vectors were added using a multiplicity of infection of 4.5:4.5:2.7 (KOS:hc-Myc:hKlf4) for *NAGLU*$^{E153K/E153K}$ skin FBs and 7:7:4.2 for *GBA1*$^{N370S/N370S}$ skin FBs. FB media with 15% (v/v) FBS was replaced 24 hours after transduction, and every 48 hours thereafter. CF1 irradiated mouse embryonic FB (iMEF; Gibco, A34180) feeder layers were plated on 0.1% (w/v) gelatin powder in dH_2O-coated plates five to six days post-transduction. Seven days post-transduction, FBs were passaged to iMEF and switched to human iPSC media, comprising of KnockOut™ D-MEM (Gibco, 10829-018) basal media with 20% (v/v) KnockOut™ Serum Replacement (Gibco, 10828-028), 1% (v/v) non-essential amino acids (Gibco, 11140-050), 1X (v/v) GlutaMAX^{TM}-I Supplement (Gibco, 35050-061), 0.4 μM (v/v) 2-Mercaptoethanol (Gibco, 21985-023), 10 ng/mL (v/v) recombinant human FGF-basic (PeproTech, 100-18B; reconstituted in Tris 0.1% (v/v) bovine serum albumin (BSA) solution (5 mM (w/v) Tris, 0.1% (w/v) BSA, pH 7.6). Reprogrammed *GBA1*$^{N370S/N370S}$ skin FBs were cultured with 0.75 mM valproic acid (w/v). Human iPSC media was replaced daily.

Human iPSCs

iPS colonies were identified emerging on iMEF based on characteristic morphology of well-defined borders enclosing tightly compact clusters of cells (Kato *et al.*, 2016; Bjørlykke *et al.*, 2019). Colonies emerging on iMEF more than 12 days post-transduction were stained with Alkaline Phosphatase Live Stain (Thermo Fisher Scientific, A14353) as per manufacturer's protocol. Reprogramming efficiency was calculated by comparing the number of emergent colonies to the number of starting skin FBs (Kogut *et al.*, 2018). Fluorescence and brightfield photomicrographs were obtained using a Leica DMI3000 B inverted microscope and QCapture™

Pro 7 Software. Colonies positive for alkaline phosphatase were scored and quartered using a 25-gauge needle and transferred to tissue culture plates coated with Vitronectin Recombinant Human Protein, Truncated (VTN-N; Gibco, A31804). iPSCs on VTN-N were cultured in Essential 8™ complete media (Gibco, A1517001) consisting of Essential 8™ basal media and 2% (v/v) Essential 8™ supplement. Media was replenished daily. iPSCs were passaged or frozen when at approximately 80% confluence. 0.1% (v/v) EDTA, pH 8.0, (Ambion, AM9260G) was used for cell dissociation as per manufacturer's instructions. Freezing media consisted of culture media with 10% (v/v) dimethyl sulfoxide (Sigma-Aldrich, D8418). iPSCs were cultured with 10 μM (v/v) Y-27632 (STEMCELL Technologies™, 72304) for 24 hours post-thawing or passaging. Normal episomally reprogrammed human iPSCs were used as controls (Thermofisher, A18944, Lot# 2191737). Cells were rinsed in phosphate buffered saline (PBS; 0.0038M (w/v) NaH_2PO_4, 0.0162 M (w/v) Na_2HPO_4, 0.008384 M (w/v) NaCl, pH 7.4) when necessitated. APLS was used periodically to identify undifferentiated colonies.

3.2.2. iPSC characterization

gDNA was isolated from skin FBs and iPSCs, PCR amplified and Sanger sequenced as per chapter 2 section 2.2.2. TaqMan™ hPSC Scorecard™ assays were performed by Thermo Fisher Scientific (Supplementary Figure 3.1).

3.2.2.1. Trilineage differentiation

AggreWell™400 6-well plates (STEMCELL Technologies, 34421) were prepared as per the manufacturer's protocol. Human iPSCs were passaged in EB medium containing KnockOut™ D-MEM/F-12 (Gibco, 12660-012) basal media with 20% (v/v) KnockOut™ Serum Replacement, 1% (v/v) non-essential amino acids, 1X (v/v) GlutaMAX^{TM}-I Supplement, and 0.1 mM (v/v) 2-Mercaptoethanol and supplemented with 10 μM (v/v) Y-27632. After 48 hours, spheroids form in individual wells. EBs are gently resuspended in EB media and centrifuged at 3,000 *g* for 30 seconds. Briefly, EBs are fixed in 4% (v/v) paraformaldehyde, rinsed with 1% (v/v) BSA in PBS and permeabilized with 0.1% Triton™ X-100. EBs are blocked with 1% BSA in PBS. EBs were simultaneously stained for ectoderm with Anti-human SOX1 NL493-Conjugated Goat IgG (R&D Systems, 967390) and for mesoderm with Anti-human Brachyury NL557-Conjugated Goat IgG (R&D Systems, 967388), or for ectoderm with Anti-human Otx-2 NL557-Conjugated Goat IgG (R&D Systems, 967389) and for endoderm with Anti-human GATA-4 NL493-Conjugated Goat IgG (R&D Systems, 967391). Stained EBs were washed once with 1% BSA in PBS, then nuclei were counterstained with 0.1 μg/mL DAPI. EBs were rinsed 3x with PBS. Bright field and fluorescence photomicrographs were obtained using a Leica DMI3000 B inverted microscope and analyzed using ImageJ v2 (Rasband, 2008).

3.2.2.2. Flow cytometry

Flow cytometry was performed on approximately 0.8-1.0 x 10^6 *NAGLU*$^{E153K/E153K}$ and *GBA1*$^{N370S/N370S}$ iPSCs isolated after the 18th and 7th passage, respectively. Isolated iPSCs were washed with PBS by centrifugation at 400 *g* for 4 minutes. The cell suspension was then fixed and permeabilized as per manufacturer's protocol (R&D Systems). IgG_1-fluorescein-conjugated (R&D

Systems, IC002F) and IgM-PerCP-conjugated (R&D Systems, IC015C) isotype controls, and Human/Mouse SSEA-4 Fluorescein-conjugated mouse IgG_3 (R&D Systems, FAB1435F) and SSEA-1-PerCP-conjugated mouse IgM (R&D Systems, FAB2155C) antibodies were used. A Guava EasyCyte HT flow cytometer (Millipore) was used for flow cytometric analysis.

3.2.3. FGF2 supplementation

MPS IIIB, GD, and normal human iPSCs were seeded at 10% density and supplemented with 0, 40 or 100 ng/mL (v/v) FGF-basic in Essential-8™ prepared media. Media and supplemental FGF-basic was replaced every 24 hours. Photomicrographs were taken at the same location on the culture well every 24 hours and analyzed for cell confluency using ImageJ v2 software (Rasband, 2008). Experiments were performed in triplicate, at minimum. Images exhibiting less than 5% or more than 30% confluence after 24 hours were omitted from analysis. Statistical analyses were performed using Kruskal-Wallis tests with post hoc Bonferroni procedure and simultaneous Mann-Whitney U paired tests, where $p \leq 0.05 = *$.

3.3. Results

3.3.1. Skin fibroblasts reprogramming

Transduced skin FBs show cytotoxicity as indicated by cell darkening and increased cell death. Cytotoxicity can be seen after three days post-SeV transduction, persisting through day seven (Figure 3.2). iPSCs appeared on iMEF between 12-30 days post-transduction for both cell lines. iPS colonies were then transferred to VTN-N for propagation (Figure 3.2). Reprogramming

efficiency was determined to be 0.2% for *GBA1*$^{N370S/N370S}$ skin FBs and 0.3% for *NAGLU*$^{E153K/E153K}$ skin FBs.

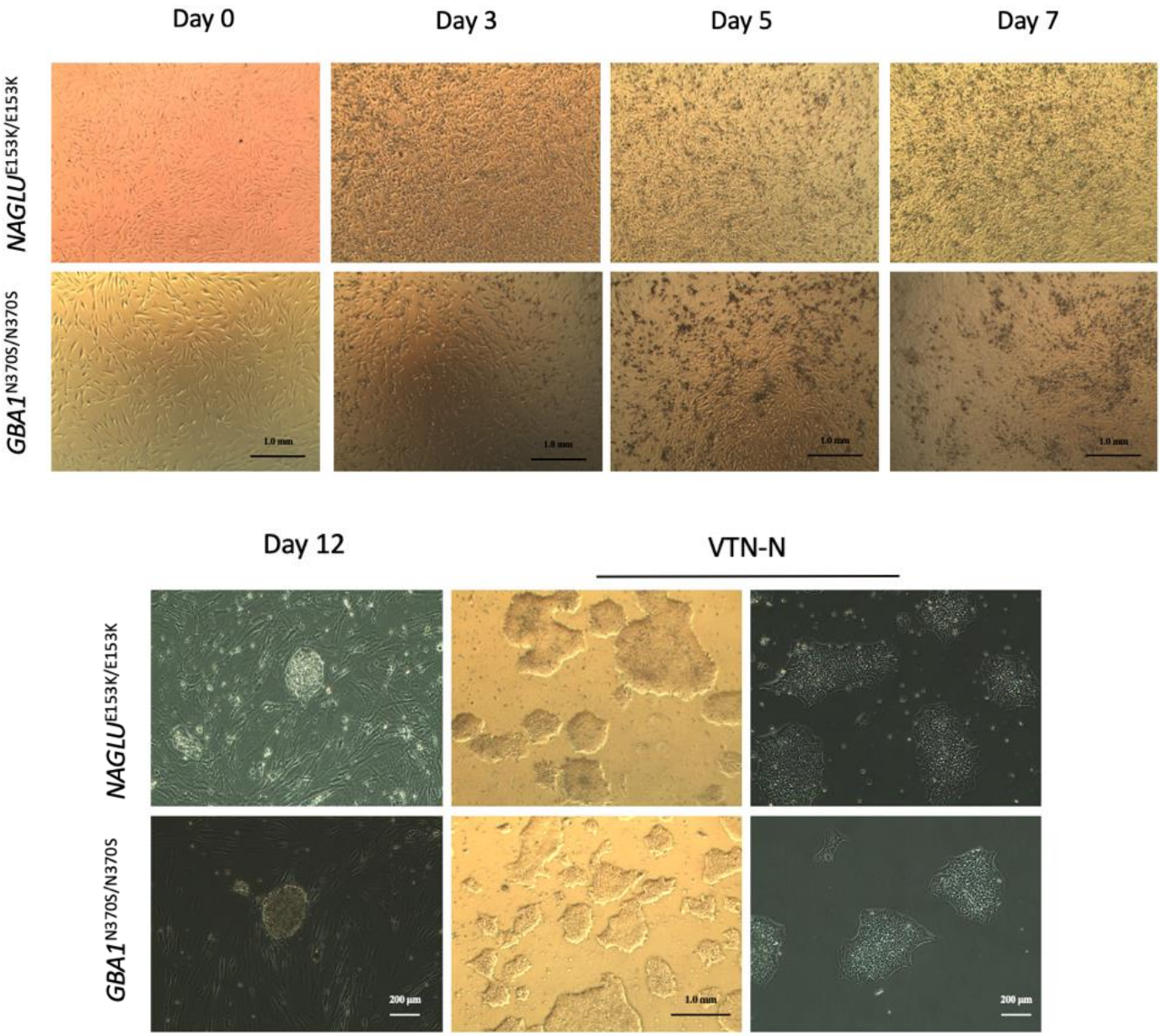


Figure 3.2 Phase contrast micrographs of GD and MPS IIIB skin FBs during reprogramming and as iPSCs emerged. Phase contrast micrographs at 4X magnification of passage 9 *NAGLU*$^{E153K/E153K}$ (top) and passage 6 *GBA1*$^{N370S/N370S}$ (bottom) skin fibroblasts during SeV transduction using CytoTune™ 2.0 with a MOI of 4.5:4.5:2.7 and 7:7:4.2, respectively. **Day**

0 Prior to transduction, **Day 3** Three days post transduction, **Day 5** Five days post transduction, **Day 7** Seven days post transduction, **Day 12** Colonies emerge on iMEF feeder layers, approximately 200 μm in diameter. Cytotoxicity can be seen after three days post-SeV transduction, persisting through day seven. iPSCs appear on iMEF between 12-30 days post-transduction. Reprogramming efficiency was determined to be 0.2% for *GBA1*$^{N370S/N370S}$ skin FBs and 0.3% for *NAGLU*$^{E153K/E153K}$ skin FBs. iPSC colonies transferred to VTN-N are shown, exhibiting typical stem cell morphology with regular borders and a high nucleus to cell ratio.

iPSCs were identified via APLS while emerging on iMEF and again after ALP+ colonies were established in feeder-free conditions (Figure 3.3A). ALP+ cells were propagated in culture, and further characterized using flow cytometry for differentiation marker SSEA-1 and pluripotency marker SSEA-4. *GBA1*$^{N370S/N370S}$ and *NAGLU*$^{E153K/E153K}$ iPSCs were isolated after the 7th and 18th passage, respectively. Flow cytometric analysis of SSEA-4 showed 77% and 78% of the iPSCs analysed were positive for SSEA-4 for *NAGLU*$^{E153K/E153K}$ and *GBA1*$^{N370S/N370S}$ iPSCs, respectively (Figure 3.3D). Flow cytometric analysis of SSEA-1 showed 5.2% and 3.7% of the iPSCs analysed were positive for SSEA-1 for *NAGLU*$^{E153K/E153K}$ and *GBA1*$^{N370S/N370S}$ iPSCs, respectively (Figure 3.3D). *NAGLU*$^{E153K/E153K}$ and *GBA1*$^{N370S/N370S}$ iPSCs were differentiated to EBs, and stained for endoderm marker GATA-4, mesoderm marker Brachyury, and ectoderm markers Otx-2 and SOX1 (Figure 3.3C). Sanger sequencing of *GBA1*$^{N370S/N370S}$ and *NAGLU*$^{E153K/E153K}$ skin FBs prior to reprogramming and iPSCs after reprogramming show the retention of the disease genotype (Figure 3.4).

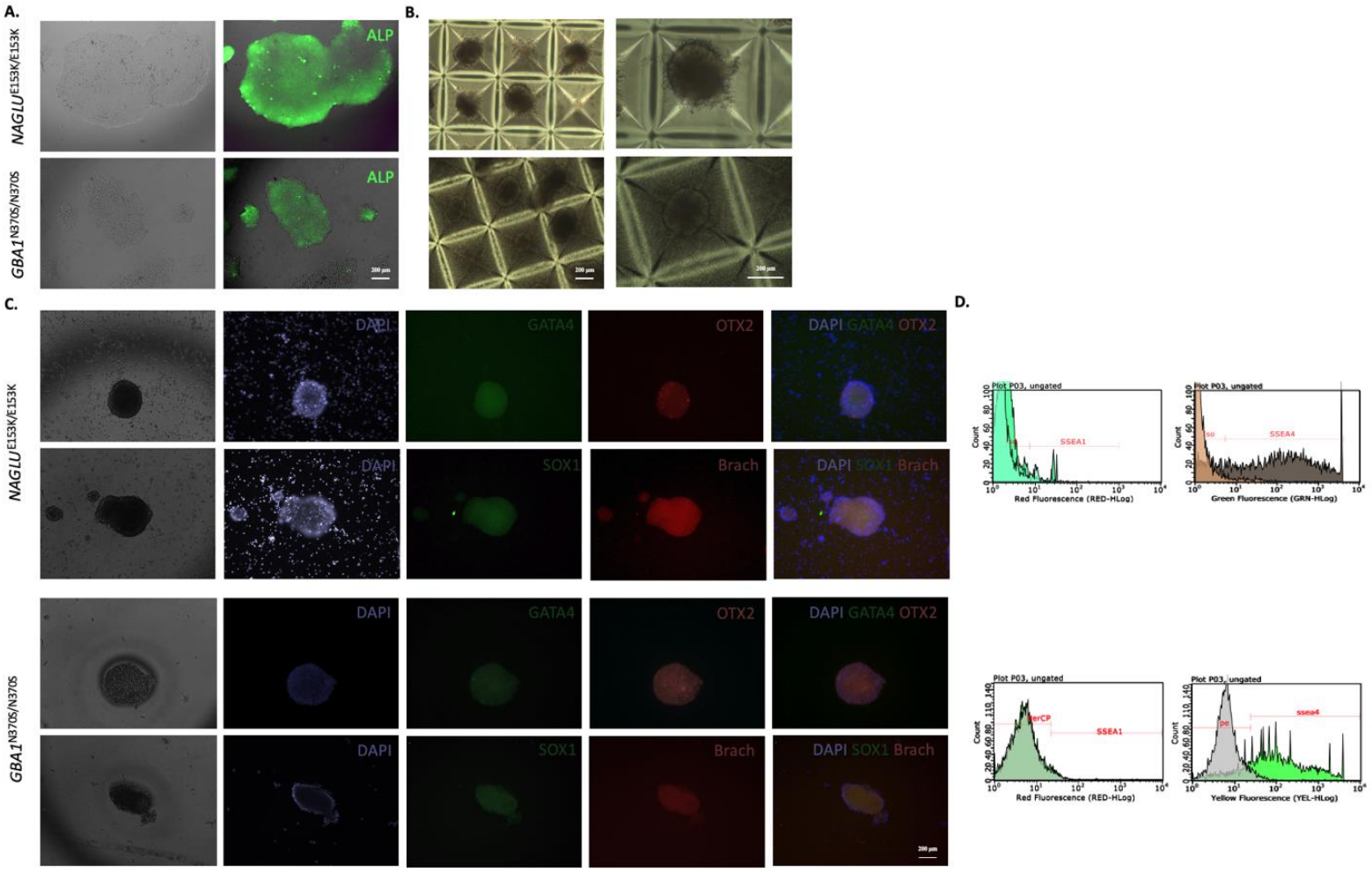


Figure 3.3 Confirmation of *NAGLU*$^{E153K/E153K}$ and *GBA1*$^{N370S/N370S}$ iPSC pluripotency. A. APLS of iPSCs plated on VTN-N **B.** EBs forming on AggreWell™ plates, exhibiting characteristic 3D structure **C.** Photomicrographs of *NAGLU*$^{E153K/E153K}$ and *GBA1*$^{N370S/N370S}$ embryoid bodies stained for endoderm marker GATA-4, mesoderm marker Brachyury, and ectoderm markers Otx-2 and SOX1. All scale bars represent 200 µm **D.** Flow cytometric analysis of *NAGLU*$^{E153K/E153K}$ and *GBA1*$^{N370S/N370S}$ after the 18th passage and 7th passage, respectively, using a guava easyCyte™ flow cytometer with the pluripotency markers SSEA-4 and differentiation marker SSEA-1. For both SSEA-4 and SSEA-1 analysis, an isotype positive control was used. Every measurement used a total of 5000 cells. Analysis of fluorescence of mouse IgG_1-fluorescein-conjugated isotype control, and positive marker SSEA-4-anti-SSEA4-fluorescein-conjucated mouse IgG_1. When 5% of the isotype control was included, 77% and 78% of the sample was positive for SSEA-4, respectively. Analysis of fluorescence of mouse IgM-PerCP isotype control, and positive marker SSEA-1-PerCP conjugated to mouse IgM (green). When 5% of the isotype control was included, 5.2% and 3.7% of the sample was positive for SSEA-1, respectively.

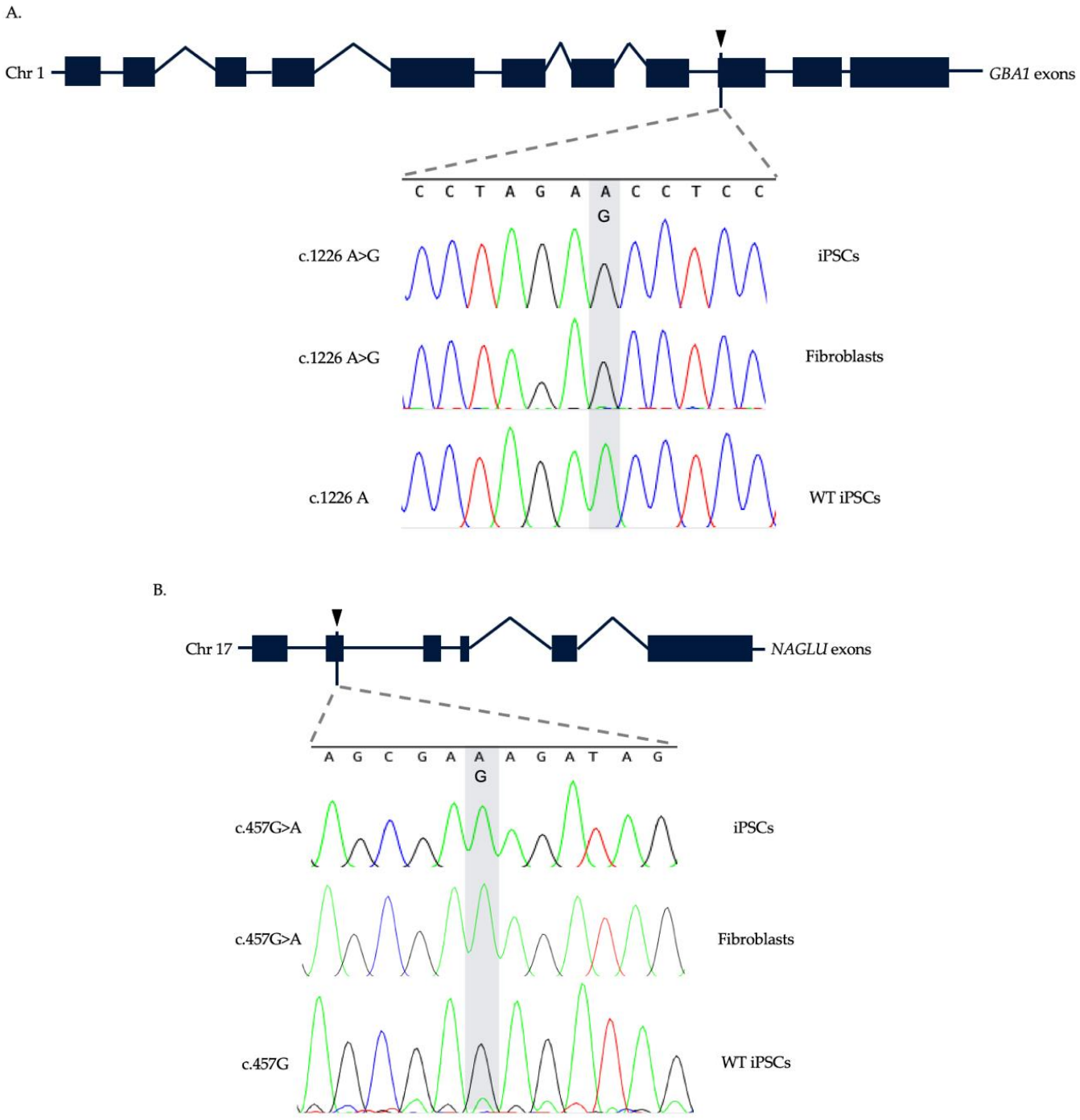


Figure 3.4 Schematic showing location of homozygous *GBA1* c.1226 A>G and *NAGLU* c.457 G>A mutations. These mutations persist throughout reprogramming to iPSCs and are absent in WT iPSCs as confirmed through Sanger sequencing. **A.** c.1226 A>G homozygous mutation is located on chromosome 1q22 in the 9th exon of *GBA1*. c.1226 is highlighted in gray, showing black

peaks designating guanine nucleotides and green peaks designating adenine nucleotides. **B.** c.457 G>A homozygous mutation is located on chromosome 17q21 in the 2nd exon of *NAGLU*. c.457 is highlighted in gray, showing green peaks designating adenine nucleotides and black peaks designating guanine nucleotides.

3.3.2. FGF2 supplementation

Experiments were conducted on passage 4-6 MPS IIIB iPSCs cultured with supplemental FGF2 over a period of 120 hrs. A significant increase in MPS IIIB iPSC confluency was observed when supplemented with 100 ng/mL for 48 hours ($p \leq 0.05$), persisting at time points 72 and 96 hours ($p \leq 0.05$). Healthy and GD control cultures were not significantly impacted by increased FGF2 concentration (Figure 3.5).

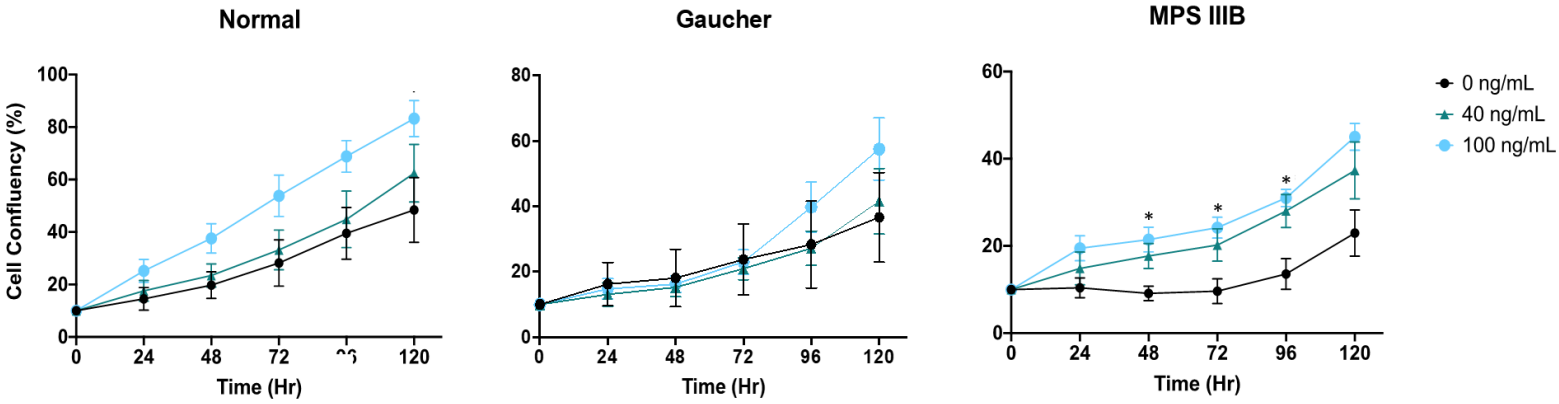


Figure 3.5 Impact of FGF2 supplementation through measurement of iPSC confluency over a period of 5 days. Three cell lines were supplemented with 0, 40 and 100 ng/mL FGF2 daily (C_f = 100, 140, and 200 ng/mL FGF2, respectively): **A.** Normal iPSCs P4-6 (n = 4), **B.** GD iPSCs P4-6 (n = 3), and **C.** MPS IIIB iPSCs P4-6 (n = 5). P-values were calculated using Kruskal-Wallis tests with post hoc Bonferroni procedure with simultaneous Mann-Whitney U paired tests (* = $p \leq 0.05$). Cells were seeded at 10% density on day zero (0 hours) and confluency was quantified every 24 hours for five days via light microscopy, followed by ImageJ v2 analysis (Rasband, 2008).

3.4. Discussion

Terminally differentiated cells are easily acquired from patients with minimally invasive methods and reprogrammed to iPSCs to provide autologous cells for disease modelling and regenerative therapies. In this chapter, we focused on the reprogramming of two skin FB cell lines to pluripotency: *NAGLU*$^{E153K/E153K}$ and *GBA1*$^{N370S/N370S}$. Mutations in *GBA1* have been found to be associated with a 4-6-fold increase in the risk of developing PD later in life (Sidransky *et al.*, 2009). The most frequent *GBA1* mutation associated with PD is the p.N370S mutation (Sidransky *et al.*, 2009). Previously, *GBA1*$^{N370S/N370S}$ skin FBs have been reprogrammed to iPSCs, although by using lentiviral methods which carries the risk of genomic integration (Panicker *et al.*, 2012; Sun *et al.*, 2015). SeV reprogramming eliminates the risk of viral integration, thus utilizing non-integrating methods is key for establishing iPSC workflows with clinical utility. To study the contribution of heterozygous *GBA1* mutations on cellular pathologies associated with PD, researchers have used SeV reprogramming methods to establish iPSC clones exhibiting the *GBA1*$^{N370S/L444P}$ genotype (Tofoli *et al.*, 2019) and *GBA1*$^{N370S/-}$ genotype (Woodard *et al.*, 2014; Rodriguez-Traver *et al.*, 2019).

Interestingly, Woodard *et al.* obtained and reprogrammed skin FBs from monozygotic twins heterozygous for p.N370S and discordant for PD (2014). They found that although both twins had approximately 50% glucocerebrosidase levels and 3-fold elevated α-synuclein levels, one twin had a more marked reduction in dopamine synthesis and release capacity and increased monoamine oxidase B expression in midbrain dopaminergic neurons, suggesting additional factors contributing to the affected twin's pathology (Woodard *et al.*, 2014). Although iPSCs have been

used to study the role of heterozygous p.N370S mutations on α-synuclein aggregation, ER stress, and mitochondrial dysfunction due to the higher prevalence of p.N370S carriers over homozygotes (Fernandes *et al.*, 2016), it would be informative to characterize the cellular impacts of homozygous p.N370S *GBA1* mutations on these cellular pathologies in addition.

The reprogramming efficiencies for *NAGLU*$^{E153K/E153K}$ and *GBA1*$^{N370S/N370S}$ of 0.3% and 0.2%, respectively, fall within the range reported in the literature. Schlaeger *et al.* compared a number of non-integrating lentiviral methods of reprogramming, showing that MKOS expressed as mRNA resulted in the highest reprogramming efficiency of 2.1%, followed by 0.077% for SeV methods, and 0.013% for episomal reprogramming (2015). Others have reported up to 1% reprogramming efficiencies using a combination of methods, including the use of small molecule aides like VPA (Huangfu *et al.*, 2008). Typical reprogramming efficiencies with SeV methods range between 0.01% and 1% (Fusaki *et al.*, 2009; Ban *et al.*, 2011; Seki *et al.*, 2010; Malik and Rao, 2013). A cell line (WG 0421) obtained from a patient with a diagnosis of MPS IIIB based on clinical and biochemical findings was previously reprogrammed to pluripotency using SeV non-integrating methods in our lab, with a reported efficiency of 0.25%, which is similar to our reported reprogramming efficiencies (Turner, 2015). Here, GD and MPS IIIB iPSCs were stained for pluripotency marker, SSEA-4, and differentiation marker, SSEA-1, to confirm pluripotency. Flow cytometric analysis of SSEA-4 yielded 77% and 78% SSEA-4-positive cells when 5% of the isotype control was included for MPS IIIB and GD iPSCs, respectively. The differentiation marker, SSEA-1, was present on 5.2% and 3.7% of cells when 5% of isotype control was included for MPS IIIB and GD iPSCs, respectively. A routinely accepted threshold for iPSCs used in

clinical applications is >70% positive for SSEA-4, thus these generated iPSCs are congruent with clinical standards (Sullivan *et al.*, 2018).

GD and MPS IIIB iPSCs were also differentiated to EBs and stained for endoderm marker GATA-4, mesoderm marker Brachyury, and ectoderm markers Otx-2 and SOX1. All four markers were expressed in EBs derived from both GD and MPS IIIB iPSCs, indicating these iPSCs possess the ability to differentiate into the three germ layers. TaqMan™ hPSC Scorecard™ qPCR assays confirmed the that these iPSCs maintain expression profiles of pluripotency (Supplementary Figure 3.1), but also reveals the sustained presence of SeV in our *GBA1*$^{N370S/N370S}$. Going forward, these *GBA1*$^{N370S/N370S}$ iPSCs will be cultured at 38°C for 3-5 days to eliminate remaining SeV (Ban *et al*., 2011). Further characterization to confirm normal karyotype should also be performed, since karyotype abnormalities, especially during extended culture, have been noted in both ESCs (Buzzard *et al.*, 2004) and iPSCs (Taapken *et al.*, 2011).

Although our reprogramming efficiency for MPS IIIB iPSCs (0.3%) is comparable with literature findings, we noticed that throughout early passages these iPSCs appeared to show increased cell death and spontaneous differentiation to 3-dimensional spheroids, akin to EBs. EBs are routinely generated from iPSCs in part by removing FGF2 from the culture media (Son *et al.*, 2011). Upon supplementing FGF2 to MPS IIIB iPS cultures and noting concomitant reduced cell death and spontaneous differentiation, we sought to measure the impact supplemental FGF2 has on MPS IIIB iPSC proliferation through the overall change in cell confluency over time with 40 and 100 ng/mL supplemental FGF2. We found that although normal control iPSCs, and GD control iPSCs appear to gain confluency over a period of 5 days with 100 ng/mL supplemental FGF2, significance is reached only with 100 ng/mL supplemental FGF2 in MPS IIIB iPS culture after 48

hours ($n = 5$, $p \leq 0.05$, Kruskal-Wallis tests with post hoc Bonferroni procedure and simultaneous Mann-Whitney U paired tests), persisting until 96 hours ($n = 5$, $p \leq 0.05$, Kruskal-Wallis tests with post hoc Bonferroni procedure and simultaneous Mann-Whitney U paired tests). The inherent mitogenic properties of FGF2 may explain the increase in cell proliferation in all three iPSC lines (Nawrocka *et al.*, 2020).

Although others simultaneously to us reprogrammed *NAGLU*$^{E153K/E153K}$ skin FBs to iPSCs, they interestingly do not report a reprogramming efficiency in this study or indicate issues with proliferation (Huang *et al.*, 2019). One key difference between Huang *et al.*'s reported reprogramming of *NAGLU*$^{E153K/E153K}$ skin FBs to iPSCs and the reprogramming procedure reported in this dissertation is that during the feeder-free stage, they cultured their iPSCs in a media containing FGF2 with higher thermostability (2019). This allows for FGF2 bioavailability to persist well beyond 24 hours unlike unmodified FGF2 which rapidly degrades at 37°C (Chen *et al.*, 2013). Going forward, culture medias with modified FGF2 with higher thermostability may be recommended for disorders of HS catabolism to ensure consistently maintained levels of FGF2 throughout the reprogramming process.

In 2011, Lemonnier *et al.* found that generation of MPS IIIB iPSCs from two patient skin FB cultures, one homozygous for a splice donor mutation c.531+1 G>C and the other homozygous for the p.R482W mutation in *NAGLU,* was less efficient than normal controls and emerging MPS IIIB colonies were smaller in diameter and failed to proliferate. This group hypothesized the accumulation of partially digested HS was impeding FGF2 signaling and chose to mitigate this issue by genetically engineering iMEF feeder layers to release NAGLU into the culture media

(Lemonnier *et al.*, 2011). They found that after several passages on iMEF-NAGLU feeder layers, iPSCs transferred to feeder free conditions and starved of NAGLU proliferate despite HS accumulation (Lemonnier *et al.*, 2011). MPS IIIB is not the only LD that has proved difficult to model using iPSCs. Huang *et al.* found that after multiple unsuccessful attempts at reprogramming skin FBs to iPSCs a lentiviral doxycycline inducible vector strategy for expressing *GAA* could be used to reduce glycogen burden on these cells (2011). Although these are viable strategies for overcoming reprogramming issues resulting from monogenic disorders with a single enzyme defect, these are not easily translatable to other diseases of HS metabolism such as hereditary multiple exostoses, age-related macular degeneration and Alzheimer's disease (Zak *et al.*, 2002; Park and Shukla, 2013; van Horssen *et al.*, 2003). In order to move iPSCs to clinical applications, xeno-free culture conditions will need to be maintained thus barring the use of murine feeder layers (Tucker *et al.*, 2012). Moreover, the addition of exogenous human recombinant lysosomal enzymes to culture media can be prohibitively expensive where stem cell culture media is already a highly costly cell culture consumable (Horiguchi *et al.*, 2018).

The role of HS as a cofactor in FGF cell signaling pathways indirectly impacts the activation of MAPK and PI 3-K/Akt pathways, crucial pathways in maintaining stem cell pluripotency (Mossahebi-Mohammadi *et al.*, 2020). Ample evidence shows that properly sulfated free and bound HS binds to FGF2, protecting this mitogen from degradation but also reducing its bioavailability (Quarto and Amalric, 1994). Furthermore, partially degraded or improperly sulfated HSPGs may fail to bind FGF2, resulting in reduced interactions at FGFRs and a reduction in FGF2 half-life (Pan *et al.*, 2005; Lemonnier *et al.*, 2011; Horiguchi *et al.*, 2018). Based on these two possible mechanisms, we propose that supplemental FGF2 will result in more FGF2-HS-

FGFR interactions either (1) through over saturation of the abundant HSPGs at the cell surface (2) or by providing more FGF2 to overcome excess FGF2 degradation.

In an attempt to address the exorbitant costs associated with iPSC culture media, a recent study examined the impact of supplemental HS to increase the half-life of FGF2 and reduce the frequency of media changes (Horiguchi *et al.*, 2018). Horiguchi *et al.* found that supplementing normal iPSC cultures with more than 5 μg/mL HS resulted in an increase of FGF2 concentration by preventing degradation (2018). Future research should focus on supplementing properly sulfated HS in iPSC models of disorders of HS catabolism to determine if increasing the half-life of FGF2 overcomes issues with proliferation. Additionally, studies have suggested that dermatan sulfate, chondroitin sulfate, and keratan sulfate may also play a role in FGF2 signalling, thus alternative iPSC culturing strategies to mitigate undesired effects of GAG accumulation should be considered when studying disorders of GAG catabolism in general (Taylor *et al.*, 2005; Sterner *et al.*, 2013). Another recent study has showed that culturing fibroblasts at hypothermic temperatures increases NAGLU enzyme activity and total enzyme levels (Meijer *et al.*, 2017). Further work should explore how hypothermic culture conditions impact NAGLU enzyme activity and total enzyme levels during skin FB reprogramming.

3.5. Conclusions

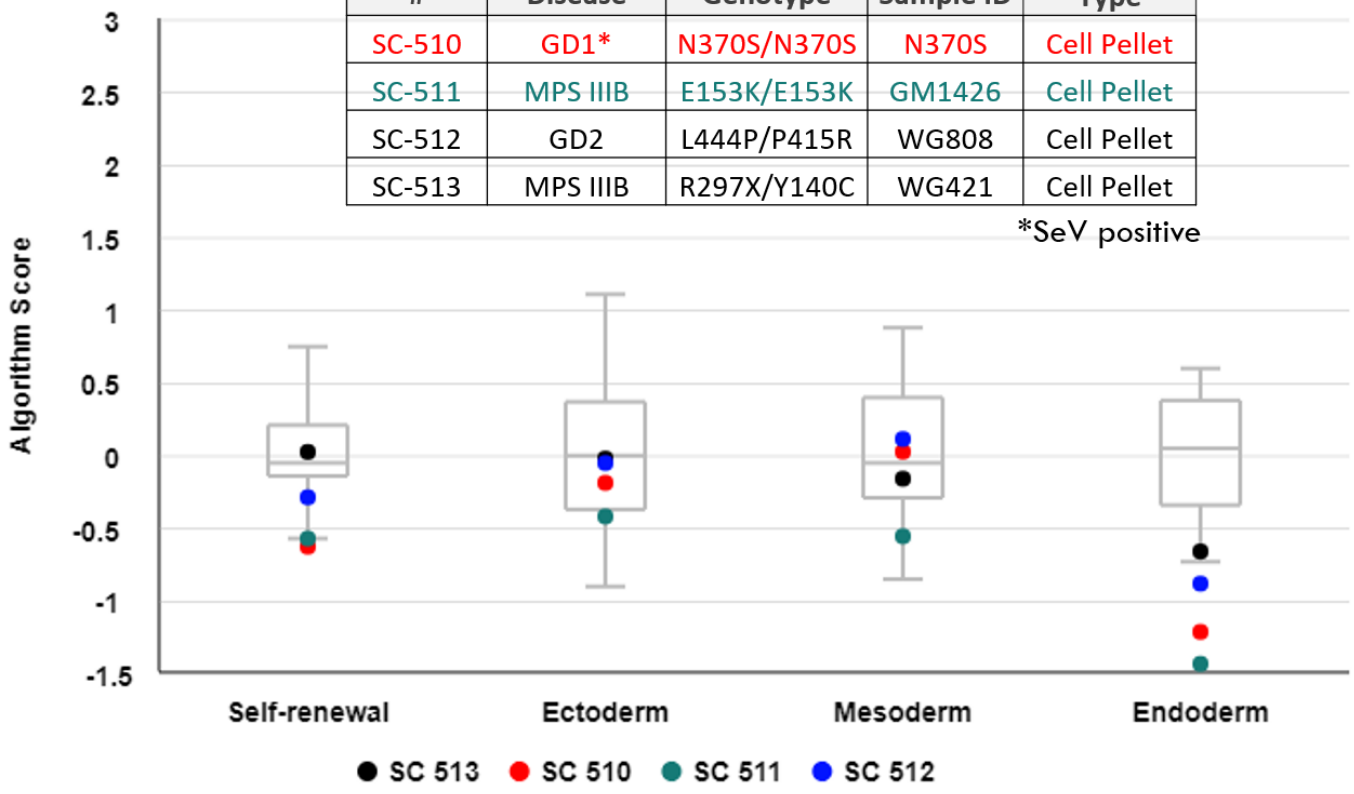
The advent of iPSC technology has progressed our ability to advance gene therapies and disease modelling. Here, we report the reprogramming of patient skin FBs derived from patients with MPS IIIB and GD with homozygous p.E153K and p.N370S genotypes, respectively. Impeded proliferation was abated with the supplementation of 100 ng/mL FGF2 in culture media for MPS IIIB iPSCs, which we have postulated to overcome cellular pathology resulting from partially degraded HS in this disorder of HS catabolism. Future work will further characterize these iPSC lines through karyotyping and investigating alternative options to overcome proliferation issues arising from HS accumulation.

3.6. Supplementary Information

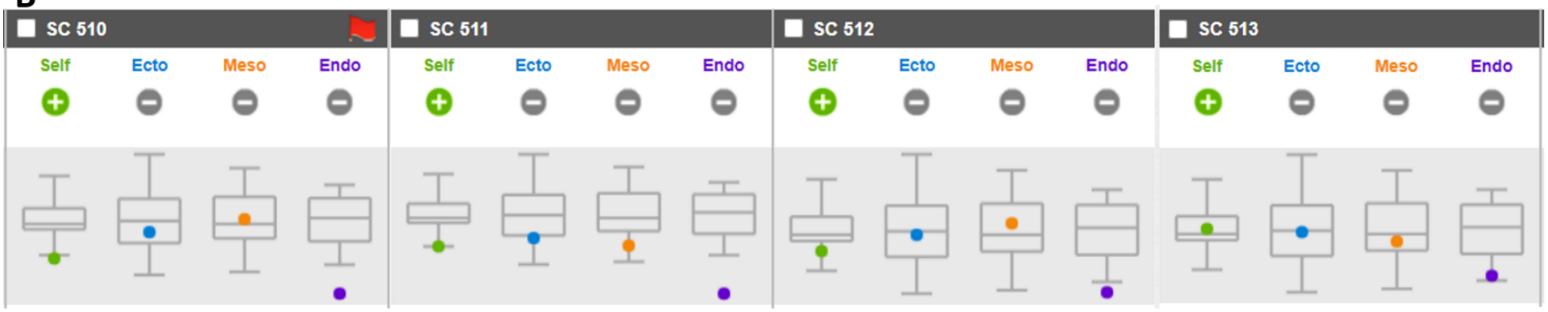
A

#	Disease	Genotype	Sample ID	Type
SC-510	GD1*	N370S/N370S	N370S	Cell Pellet
SC-511	MPS IIIB	E153K/E153K	GM1426	Cell Pellet
SC-512	GD2	L444P/P415R	WG808	Cell Pellet
SC-513	MPS IIIB	R297X/Y140C	WG421	Cell Pellet

*SeV positive



B



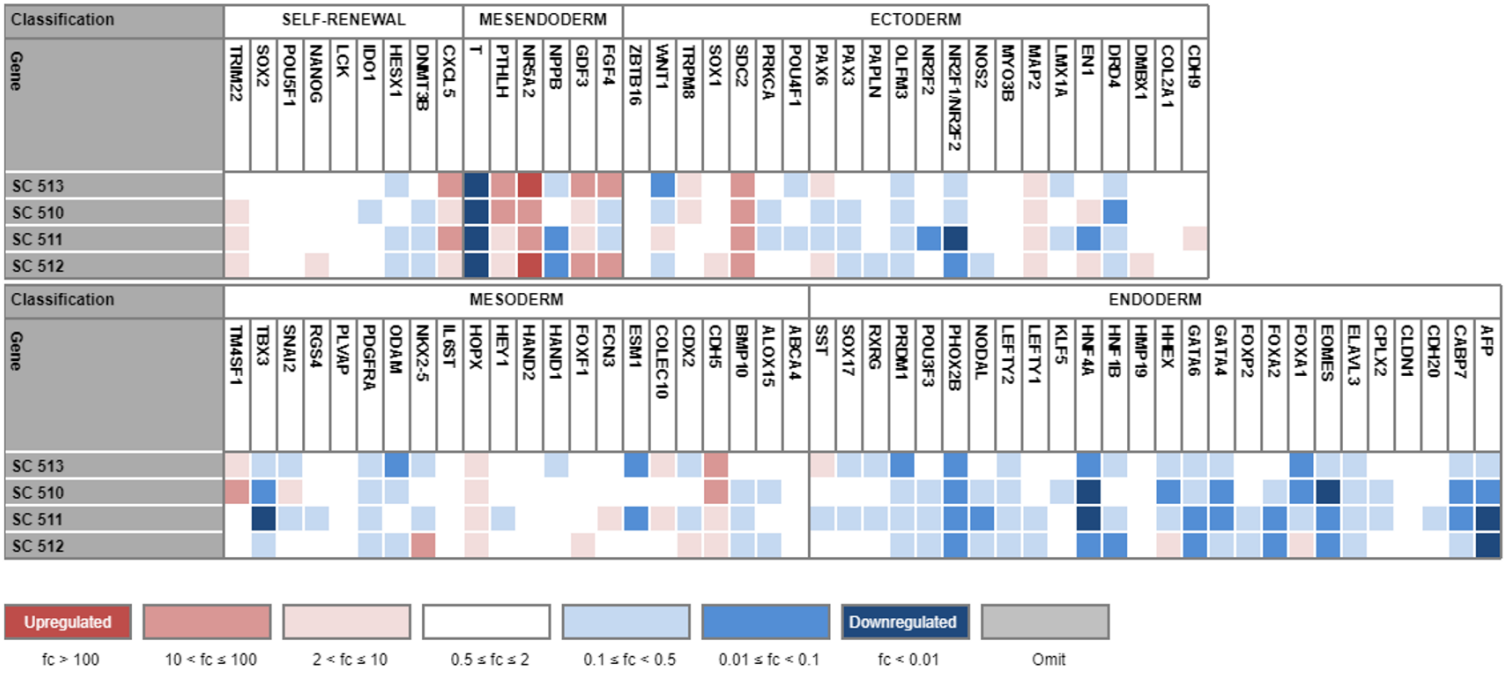
C

Sample Name	Self-renewal	Ectoderm	Mesoderm	Endoderm
SC 513	0.02	-0.03	-0.17	-0.67
SC 510	-0.64	-0.20	0.02	-1.22
SC 511	-0.58	-0.43	-0.56	-1.44
SC 512	-0.30	-0.06	0.10	-0.89

Gene expression relative to the reference standard

Upregulated						Downregulated
x > 1.5	1.0 < x ≤ 1.5	0.5 < x ≤ 1.0	-0.5 ≤ x ≤ 0.5	-1.0 ≤ x < -0.5	-1.5 ≤ x < -1.0	x < -1.5

D



Supplementary Figure 3.1. TaqMan™ hPSC Scorecard™ qRT-PCR assays for expression of markers for pluripotency and differentiation. TaqMan™ hPSC Scorecard™ qRT-PCR assays (Thermo Fisher Scientific) were performed on two MPS IIIB (*NAGLU*$^{E153K/E153K}$, sample SC-511 and *NAGLU*$^{Y140C/R297X}$, sample SC-513), one GD1 (*GBA1*$^{N370S/N370S}$, sample SC-510), and one GD2 (*GBA1*$^{L444P/P415R}$, sample SC-512) hiPSC cell lines that were reprogrammed with SeV. **A.** Box plots representing a range of scores for an undifferentiated reference set compared to sample scores (plotted in colour) assessed for markers of self-renewal, ecto-, meso-, and endoderm. **B.** Data from (**A.**) separated based on individual sample. Red flag indicates the presence of SeV detected. **C.** Overall gene expression profiles represented as upregulated (red) or downregulated (blue) in comparison to an undifferentiated reference set. **D.** Specific gene expression profiles represented as upregulated (red) or downregulated (blue) in comparison to an undifferentiated reference set. Data was generated and analysed by Thermo Fisher Scientific.

4. Genome editing outcomes in *GBA1* and *NAGLU*

4.1. Introduction

Genome editing systems are of particular interest when developing therapeutic strategies for rare LDs. Lysosomal hydrolases, which are non-functional in LDs, are conveniently tagged with M6P residues, allowing for trafficking of lysosomal hydrolases to the lysosome via M6P receptors. Fortuitously, M6P receptors are also embedded in the plasma membrane of mammalian cells. This allows for the passage of lysosomal hydrolases between neighbouring cells in a process called “cross-correction” (Fratantoni *et al.*, 1968). For this reason, and since only slight increases in total cellular lysosomal hydrolase has been shown to result in attenuated disease (Leinekugel *et al.*, 1992; Schueler *et al.*, 2004), *in-* or *ex vivo* genome editing strategies are particularly promising for LDs. Of note, genome editors have been used to install *IDUA* and *IDS* at the albumin locus, a safe harbour for *in vivo* gene insertion in hepatocytes under the liver-specific promoter human α-1-antitrypsin along with the apolipoprotein E enhancer (Ou *et al.*, 2019, Laoharawee *et al.*, 2018). This method has progressed to phase I/II clinical trials with interim results indicating small increases in plasma IDS and IDUA activity in patients (NCT02702115; NCT03041324).

Similarly, progress has been made in establishing a pre-clinical model for GD gene therapy by knocking in *GBA1* to the CCR5 safe harbour locus (Scharenberg *et al.*, 2020). However, knocking in a gene instead of repairing the existing mutant gene may not mitigate all the disease symptoms. Patients with the most severe form of GD die in infancy, but those who have the milder from (type I; GD1) or who are carriers for *GBA1* mutations have a 4 to 6-fold increased risk in developing PD later in life (Sidransky *et al.*, 2009; Bultron *et al.*, 2010). This increased risk in

developing PD is associated with the aggregation of mutant GCase with the aggregate protein α-synuclein (Mazzulli *et al.*, 2011; Sun *et al.*, 2015; Xu *et al.*, 2011; Xu *et al.*, 2014; Sardi *et al.*, 2011). Protein misfolding is well known to cause amyloidogenic disorders (Hashimoto *et al.*, 2003; Chiti and Dobson, 2017). The impact of protein misfolding, mitochondrial dysfunction, and oxidative stress will require further study to fully elucidate the long-term impact of mutant lysosomal hydrolases in amyloidogenic disorders. Given these complications, it is important to explore alternative genome editing strategies that target the mutant gene of interest.

Since 2012 when an acquired bacterial immune system bioengineered to target DNA with precision was proposed as a programmable RNA-guided genome editing tool, myriad genome editing strategies have been explored in the context of disease modelling and therapeutic development (Jinek *et al.*, 2012). The preeminent RNA-guided genome editing system is CRISPR-Cas9 and has since overtaken the fields of microbiology, molecular biology, genetics, and medicine, to name but a few. CRISPR-Cas9 can be used to modify DNA sequence *in*- and *ex vivo* through the use of a short 17-23 nucleotide gRNA molecule that has sequence complementarity to a genomic site of interest. The 3’ end of this RNA fits into a binding pocket within the endonuclease Cas9, guiding Cas9 to the genomic site of interest. Cas9 endonuclease creates a DSB three nucleotides upstream from a PAM recognition sequence, 5’NGG3’. Upon cleaving, one of two DNA repair pathways, NHEJ or HDR, will repair the DSB. NHEJ leads to indels at the site of the break. This pathway is commonly leveraged to knock-out genes, thus being a suitable method for creating cellular and pre-clinical models of disease (Latour *et al.*, 2019; Zhang *et al.*, 2017), but is also useful for determining on-target efficiency of gRNAs prior to directing more precise DNA repair (Sentmanat *et al.*, 2018). HDR is a templated pathway, making use of a region of

sequence similarity to direct precise repair at the break site (Pardo *et al.*, 2009). HDR can be exploited to introduce precise changes at the site of the break and has been shown extensively to lead to single nucleotide changes, precise introduction of desired insertions or deletions and otherwise installation of edits within the region of the cut site. These applications are discussed in further detail in chapter 1 and in Christensen *et al.* (2019). In this chapter, we describe outcomes of the CRISPR-Cas9 system to target mutations in *NAGLU* (c.457 G>A, c.889 C>T, and c.1073 C>T) and *GBA1* (c.1226 A>G) *in vitro*.

HDR is often less efficient than desired for introducing genetic changes at a site of interest. Even in transformed easy-to-transfect cell lines, HDR is often observed in less than 10% of the total cell population (Anzalone *et al.*, 2019; Maruyama *et al.*, 2015; Mali *et al.*, 2013). This is due in part to cell cycle synchrony requirements for editing (Chang *et al.*, 2017; Yeh *et al.*, 2018; Lin *et al.*, 2014; Heyer *et al.*, 2010) and to the activation of p53 leading to apoptosis in response to DSBs (Haapaniemi *et al.*, 2018; De Zio *et al.*, 2013). Additionally, with CRISPR HDR, off-target mutations and on-target indels are often more common than the intended edit (Anzalone *et al.*, 2019). To mitigate these inherent issues with the CRISPR-Cas9 system, two alternative genome editing systems have been bioengineered: base editors and prime editors.

The base editing system was proposed as an alternative genome editing strategy to correct point mutations without introducing DSBs (Komor *et al.*, 2016). nCas9 linked to modified RNA deaminases, APOBEC1 and ecTadA, result in adenine or cytosine deamination within a predictable five-nucleotide window (Komor *et al.*, 2016; Gaudelli *et al.*, 2017; Figure 4.1). Base editors reduce off-target risk and increase on-target efficiency due in part to the use of a nCas9

that creates only a single phosphodiester bond nick in the DNA backbone instead of a DSB (Gaudelli *et al.*, 2017). Base editors are an improvement to CRISPR-HDR with respect to efficiency, but they are only capable of targeting adenine to be converted to inosine (guanine) or cytosine to be converted to thymine within five-nucleotide base editing windows that are determined by the presence of a nearby PAM. Transversions (C-to-G, A-to-T, A-to-C, and T-to-G) cannot be attained using the base editing systems. Additionally, the presence of unintended target nucleotides within the five-nucleotide base editing window (referred to as bystander bases) may result in undesired edits in addition to or instead of the target DNA edit. Details of the benefits and pitfalls of base editing systems is described at length in Christensen *et al.* (2019).

Given these aforementioned drawbacks to base editing systems, a hybrid form of CRISPR HDR and base editing called "prime editing" was developed (Anzalone *et al.*, 2019). Prime editors use a nCas9 linked to an RT. The gRNA used with the prime editing system (referred to as a pegRNA) contains a 5' spacer sequence that is complementary to 20 nucleotides of the target genomic site, a 3' extender sequence that is comprised of a primer binding site at the terminal 3' end and is complementary to the opposite strand at the target site, and an RT template immediately 5' of the primer binding site that acts as a template for RT. Linking these 3' and 5' pegRNA regions is an invariable scaffold RNA molecule that interacts with a binding pocket within nCas9. This scaffold is identical to the 3' terminal region of a simple gRNA, such as those used with the base editing and CRISPR-Cas9 systems. The RT linked to nCas9 utilizes the RT template within the pegRNA to direct reverse transcription. RT templates are designed to include the desired edit at the DNA target site (Anzalone *et al.*, 2019; Figure 4.1).

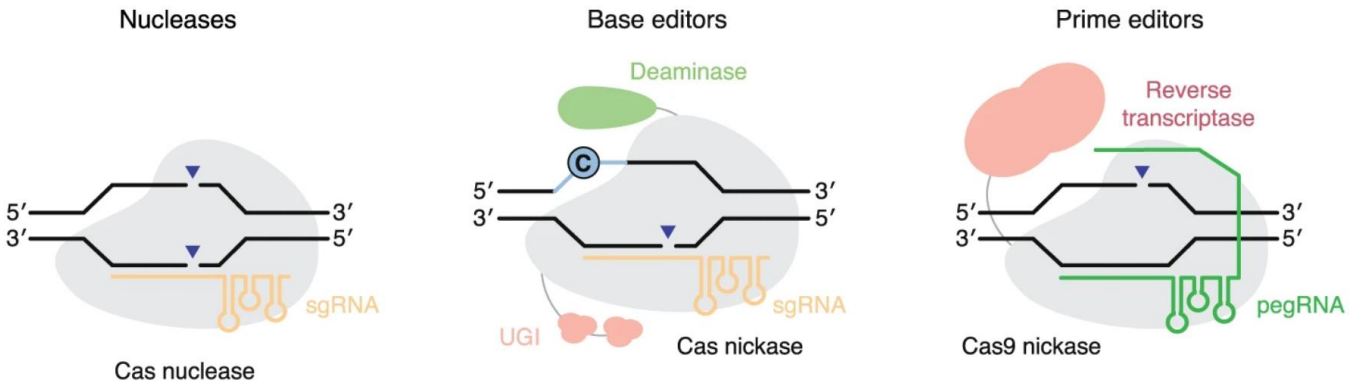


Figure 4.1. Schematic diagram of three RNA-guided endonuclease genome editing systems. Cas nucleases (left) are guided to a target site of interest by a sgRNA (yellow) creating a DSB at a predictable site three nucleotides upstream from a PAM. Base editors (middle) are similarly guided to a target site of interest by a sgRNA (yellow), but contain a mutant Cas nickase (nCas) with catalytic activity only in the HNH domain and not the RuvC domain (Komor *et al.*, 2016; Gaudelli *et al.*, 2017). A deaminase (green) is linked to nCas through a linker, capable of deaminating within a five-nucleotide base editing window 13-17 nucleotides upstream from the PAM. Depicted here (middle) is a CBE, which requires UGIs to prevent the reversion of uracil to cytosine during the transition to thymine (Komor *et al.*, 2016). Prime editors (right) make use of an extended sgRNA (pegRNA, green) which possesses a 3' extender and primer binding sequence that acts as a RT (pink) template and RT primer, respectively. Modified from Anzalone *et al.*, 2020.

Base and prime editors are more efficient at generating intended edits at a DNA target site of interest at most gDNA sites tested (Anzalone *et al.*, 2019; Haapaniemi *et al.*, 2018; Koblan *et al.*, 2018; Komor *et al.*, 2016). Prime editors allow for more robust changes at a target site, including transversions and small insertions or deletions. The efficiency of base and prime editing depends heavily on the characteristics of the target site. Since base and prime editing systems offer alternative and potentially more efficient mechanisms to elicit point mutation correction, the genome editing outcomes using these genome editing strategies to target the common c.1226 A>G mutation in *GBA1* are discussed here.

4.2. Materials and methods

4.2.1. General cell culture

HEK293 and fibroblast culture

HEK293 cells was a generous gift from Dr. Juan Ausio, University of Victoria. MPS IIIB *NAGLU*$^{-/-}$skin FBs GM01426 and GM02931 were obtained from Coriell Biorepository, and WG0421 were obtained from McGill Biorepository. GD *GBA1*$^{N370S/N370S}$ skin FBs was a generous gift from Dr. Mia Horowitz, Tel Aviv University. HEK293 and FBs were cultured in Dulbecco's modified eagle medium (Gibco, 11960-044) supplemented with 10% (v/v) FBS (ThermoFisher, 16000-044) and 1% (v/v) non-essential amino acids (Gibco, 11140-050) at 37°C/5% CO_2. Media was replenished every 3-4 days and cells were passaged or frozen at 75-85% confluence. FBs were disassociated using 0.025% Trypsin-EDTA (Gibco, 154000-054). HEK293 cells were disassociated mechanically. Freezing media consisted of culture media with 10% (v/v) dimethyl sulfoxide (Sigma Aldrich, D8418).

Human iPSCs

iPSCs were reprogrammed to pluripotency using SeV non-integrating reprogramming methods (see chapter 3 for details). iPSCs were cultured in Essential 8™ complete media (Gibco, A1517001) consisting of Essential 8™ basal media and 2% (v/v) Essential 8™ supplement on tissue culture plates coated with Vitronectin (VTN-N) Recombinant Human Protein, Truncated (Gibco, A31804) as per the manufacturer's instructions. Media was replenished daily. iPSCs were passaged or frozen when approximately 80% confluent. 0.1% (v/v) EDTA, pH 8.0, (Ambion, AM9260G) was used for cell dissociation as per manufacturer's instructions. Freezing media consisted of culture media with 10% (v/v) dimethyl sulfoxide (Sigma-Aldrich, D8418). iPSCs were cultured with 10 μM (v/v) Y-27632 (STEMCELL Technologies™, 72304) for 24 hours post-thawing or passaging. Normal episomally reprogrammed human iPSCs were used as controls (Thermofisher, A18944, Lot# 2191737). Cells were rinsed in PBS (0.0038M (w/v) NaH_2PO_4, 0.0162 M (w/v) Na_2HPO_4, 0.008384 M (w/v) NaCl, pH 7.4) when necessitated. MPS IIIB iPSCs were cultured in the presence of 100 ng/mL supplemental FGF2.

4.2.2. gRNA design and validation

gRNA design

CHOPCHOP v2 software was used to design Cas9 gRNAs targeting *GBA1* c.1226 A>G, and *NAGLU* c.457 G>A, c.889 C>T, and c.1073 C>T (Labun *et al.*, 2019; Montague *et al.*, 2014; Labun *et al.*, 2016). Candidate gRNAs were scored and selected based on GC content, 4-nucleotide self-complementarity stem, and off-target abundance and severity (Cong *et al.*, 2013; Montague

et al., 2014; Labun *et al.*, 2016). gRNAs were further analysed for proximity to the target mutation, suitable orientation for optimal HDR, and for the ability to silently mutate the PAM site with an appropriate HDR template. Base editing gRNAs were designed in Benchling (Biology Software, 2020). pegRNAs were designed in pegFinder (Chow *et al.*, 2020).

RNA-guided endonuclease-restriction fragment length polymorphism analysis

gDNA was amplified with primers flanking the Cas9 target site (see 4.2.5 for details). PCR products were run on a 2% agarose gel in 1x Tris/Borate/EDTA running buffer. DNA concentration was calculated using densitometry. Nuclease Free Duplex Buffer (Integrated DNA Technologies, 11-04-02-01) containing 5 μM Alt-R® CRISPR-Cas9 tracrRNA (Integrated DNA Technologies, 1072532) and 5 μM crRNA was incubated for 5 minutes at 95°C, then cooled to room temperature. NEBuffer™ 3.1 (New England Biolabs, B7203S) containing 1 μg Alt-R™ S.p. Cas9 nuclease 3NLS (Integrated DNA Technologies, 1074181) and 2.2 μM complexed tracrRNA/crRNA was brought to 10 μL total volume in dH_2O and incubated at room temperature for 5-10 minutes. 125 ng DNA template was added to this reaction mixture and incubated at 37°C for 1 hour. Cas9 nuclease was inactivated with 3 μg RNase A (Boehringer Mannheim, 109 169) at 37°C for 15 minutes, followed by the addition of 1.7 mg/mL of proteinase K (Ambion, 25530-015) and incubated at 58°C for 15 minutes. Digest products were run on a 2% agarose gel in 1x Tris/Borate/EDTA running buffer and stained using 0.5 μg/mL Ethidium Bromide (Sigma-Aldrich, E7637). Cleavage efficiency was calculated using ImageJ v2 software (Rasband, 2018) and densitometry to compare the proportion of cleaved products to the uncleaved band.

T7 endonuclease I analysis

gDNA was amplified with primers flanking the Cas9 target site (see 4.2.5 for details). The GeneArt™ Genomic Cleavage Detection Kit (Thermo Fisher Scientific, A24372) positive control template was amplified as per the manufacturer's protocol. PCR products were run on a 2% agarose gel in 1x Tris/Borate/EDTA running buffer. DNA concentration was calculated using densitometry. 150 ng of DNA was brought to a total of 10 µL in dH_2O and 1x NEBuffer™ 2 (New England Biolabs, B7202). These samples were denatured then hybridized on a thermal cycler using the following protocol: 95°C for 10 min, 85°C for 1 min, 75°C for 1 min, 65°C for 1 min, 55°C for 1 min 45°C for 1 min, 35°C for 1 min, 25°C for 1 min, 4°C hold. Ten units of T7 endonuclease I (T7EI; New England Biolabs, E3321) was added to the reaction mixtures and incubated at 37°C for 1 hour. Reactions were inactivated by adding 1.7 mg/mL of proteinase K (Ambion, 25530-015) and incubating for 5 minutes at 37°C. Digest products were run on either 2% agarose gel or 8% polyacrylamide gel in 1x Tris/Borate/EDTA running buffer and stained using 0.5 µg/mL Ethidium Bromide (Sigma-Aldrich, E7637). Cleavage efficiency was calculated using ImageJ v2 software (Rasband, 2018) and densitometry to compare the proportion of cleaved products to the uncleaved band. Editing efficiency is calculated using the following calculation:

$$\% \text{ Editing} = 100 \times [1- (1 - \text{fraction cleaved})^{1/2})]$$

(Guschin *et al.*, 2010)

4.2.3. Plasmid construction and bacterial transformations

The following plasmids were generous gifts from Dr. David Liu (Harvard University): pU6-pegRNA-GG-acceptor (Addgene plasmid #132777), pCMV-PE2 (Addgene plasmid #132775), pCMV-PE2-P2A-GFP (Addgene plasmid #132776), pU6-Sp-pegRNA-HEK3_CTT_ins (Addgene plasmid #132778), pCMV_ABEmax_P2A_GFP (Addgene plasmid #112101), pCMV_BE4max_P2A_GFP (Addgene plasmid #112099). The following plasmid was a generous gift from Dr. Keith Joung (Harvard University): MLM3636 (Addgene plasmid #43860). The following plasmid was a generous gift from Dr. Raad Nashmi (University of Victoria): pVenus/CS2. The following ligation and transformation protocol was adapted from Kwart *et al.*, 2017: DNA oligonucleotides were annealed in Nuclease-Free Duplex Buffer (Integrated DNA Technologies, 11-01-03-01) at a concentration of 2 nM and annealed by thermal cycling at 95°C for 5 minutes, 85°C for 1 minute, 75°C for 1 minute, 65°C for 1 minute, 55°C for 1 minute, 45°C for 1 minute, 35°C for 1 minute, 25°C for 1 minute and hold at 4°C. sgRNA scaffolding vector (MLM3636) was linearized using 0.1 μg/μL expression vector, 1x NEBuffer™ 3.1, and 5 units of *Bsm*BI (New England Biolabs, R0739) to a total volume of 50 μL. The reaction was incubated for 1 hour at 37°C. Digested products are PCR purified and confirmed by gel electrophoresis to be 2.7 kb in size. 37.5 ng *Bsm*BI-digested vector backbone was combined with the annealed DNA oligonucleotides at 0.6 μM final concentration in NEB Quick ligation buffer with 200 units of T4 DNA ligase (New England Biolabs, M0202S) and incubated at room temperature for 10 minutes, followed by heat inactivation at 65°C for 10 minutes. Subcloning Efficiency™ DH5α Competent Cells (Thermo Fisher Scientific, 18265017) were cultured for 14 hours shaking at 300 RPM in a controlled environment incubator shaker (New Brunswick

Scientific Co. Inc) at 37°C. A 50 μL aliquot of Subcloning Efficiency™ DH5α Competent Cells was then incubated on ice for > 40 minutes. Competent cells were centrifuged at 6,570 *g* for 5 minutes at 4°C. Cells were resuspended in 100 μL ice cold transformation buffer (10 mM MOPS, 15 mM $CaCl_2$:$2H_2O$, 250 mM KCl, 10.88 g/L $MnCl_2$:$4H_2O$ pH 6.7) with 5 μL DMSO and incubated on ice for 10 minutes. Ligation reaction mixture was added to transformation mixture and incubated on ice for 20 minutes. Cells were heat shocked at 42°C for 45 seconds, then returned to ice for 5 minutes, then recovered in SOC media (0.5% yeast extract, 2% tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, and 20 mM glucose) at 37°C for 1 hour. Cells were centrifuged at 8,200 *g* for 4 minutes and resuspended in 150 μL warmed SOC media. Transformed *E. coli* were spread on agar plates containing 100 μg/mL ampicillin (Sigma-Aldrich, A0166). Agar plates were incubated at 37°C for 24-48 hours. *E. coli* colonies are then isolated, prepped and Sanger sequenced to confirm correct ligated sequence (Supplementary Figures 4.1 – .2).

For pegRNA vector preparation, Golden Gate Assemblies were performed as modified from Anzalone *et al.* (2019) using the NEB® Golden Gate Assembly Kit (BsaI-HF®v2; New England Biolabs, E1602) with three components: 5’ pegRNA spacer annealed oligonucleotides, pegRNA invariable scaffold oligonucleotides, and 3’ pegRNA extender/primer binding site annealed oligonucleotides. Subcloning Efficiency™ DH5α Competent Cells were transformed as previously described. Transformed colonies with ligated pegRNA sequence lose a red fluorescent protein cassette and can be visually identified and isolated (Supplementary Figure 4.4). Accurate ligation of pegRNA oligonucleotide components was confirmed via Sanger sequencing (Supplementary Figure 4.3). QIAprep Spin Miniprep Kits (QIAGEN, 27104), ZymoPURE Plasmid Midiprep Kits (Zymo Research, D4208T), and QIAGEN® Plasmid Maxi Kits (QIAGEN,

12162) were used to purify plasmids as per manufacturer's protocol. Endotoxin removal was performed.

4.2.4. Mammalian cell transfection

Lipofection

Spacer (crRNA) RNA molecules were synthesized in 2 nmol amounts (Integrated DNA Technologies) and resuspended to 100 μM in IDTE (Integrated DNA Technologies, 11-01-02-02). Alt-R™ CRISPR-Cas9 tracrRNA 5' ATTO™ 550 were synthesized in 5 nmol amounts (Integrated DNA Technologies, 1075934) and resuspended to 100 μM in IDTE. RNP complexes were formed by combining the following: 1 ng/μL Alt-R™ S.p. Cas9 nuclease 3NLS (Integrated DNA Technologies, 1074181), 0.25 ng/μL crRNA/tracrRNA, and 1.3% v/v Lipofectamine™ Cas9 Plus™ Reagent in OptiMEM™ I reduced serum medium (Thermo Fisher Scientific, 31985062; adapted from Yu *et al.,* 2016). For CRISPR HDR, ssODN correction templates were added to the reaction mixture at a final concentration of 4 ng/μL. RNP complexes were added to 6% Lipofectamine™ CRISPRMAX™ Cas9 transfection reagent (Thermo Fisher Scientific, CMAX00008) as per manufacturer's protocols. Cultured human iPSCs were transfected at 70% confluence, then passaged and reverse transfected (Yu *et al.*, 2016). Transfected iPSCs were incubated for 18 hours at 37°C/5% CO_2. Culture medium was removed and replaced with FBM. Cells were lysed 48 hours post-transfection with QuickExtract™ DNA Extraction Solution (Lucigen, QE09050) as per manufacturer's protocol. Cell lysate was used directly as template for PCR. RNP complex and Lipofectamine™ CRISPRMAX™ Cas9 transfection reagent

concentrations were optimized using Alt-R® Human *HPRT1* Positive control crRNA, 2 nmol (Integrated DNA Technologies, 1072551) followed by T7EI assays.

Electroporation

Electroporation conditions were optimized based on previously described parameters (Table 4.5; Thermo Fisher Scientific) for human iPSCs. Cells are grown to 70-80% confluent, then passaged. Cells were rinsed with PBS twice, passaged and centrifuged at 2,100 *g* for 5 minutes at 4°C, resuspended in PBS and centrifuged again at 2,100 *g* for 5 minutes at 4°C. Prime or base editor plasmids are combined with sgRNA plasmids or pegRNA plasmids at a 3:1 ratio (750 ng: 250 ng) and electroporated into 3.0 – 5.5 x 10^6 cells at 1200V/ 20 ms/ 2 pulses in a total volume of 10 μL, then plated on 24 well plates. Cells were electroporated using the NEON™ Transfection System (Thermo Fisher Scientific) according to the manufacturer's protocol. pVenus/CS2 plasmids were used as a positive control. Media was changed 24 hours post-electroporation and cells were cultured for 96 hours prior to lysis for downstream analysis, unless otherwise specified.

4.2.5. PCR

gDNA was isolated using either the DNeasy® Blood & Tissue kit (QIAGEN, 69504), QuickExtract™ DNA Extraction Solution (Lucigen, QE09050), or Extracta DNA Prep for PCR – Tissue (Quantabio, 95091-002) as per manufacturer's protocol. DNA amplification was carried out by PCR using Platinum Green Hot Start PCR Master Mix (Thermo Fisher Scientific, 13001013) with GC Enhancer as per manufacturer's protocol. DNA oligonucleotides were

obtained from Integrated DNA Technologies. Primers were used at a final concentration of 600 nM unless otherwise stated. All PCR primer oligonucleotide sequences and thermal cycling conditions are listed in Supplementary Table 4.1. When necessary, PCR products were purified using the QIAquick PCR Purification Kit (QIAGEN, 28104) as per manufacturer's protocol.

4.2.6. Screening assays and statistical analyses

Mismatch PCR primers were designed and used to create a restriction site for *Hpy*CH4III and *Bco*DI for identification of c.889 C>T and c.457 G (New England Biolabs, R0618S). These protocols and digest parameters are described in Christensen *et al.* (2017) and chapter 2. Purified DNA amplicons were Sanger sequenced by Eurofins Genomics (with TIDE protocol, where applicable). Raw DNA sequencing data was subjected to Inference of CRISPR Edits (ICE) analysis (Synthego) or Tracking of Indels through Decomposition (TIDE), where indicated. Experiments were conducted in triplicate, where possible. Means and standard errors were calculated. P values were calculated using student's t-test with two tails and heteroscedasticity. Assumption of normality was made, but q-q plots were not generated or analysed, where applicable (*p val <0.05, **p val <0.01).

4.3. Results

4.3.1. gRNA design and validation

4.3.1.1. gRNA design and validation for CRISPR HDR

gRNAs were designed to target three pathogenic mutations in *NAGLU* and one pathogenic mutation in *GBA1*. Considerations for gRNA design included mutation prevalence, predicted disease severity, cut site proximity to target base, GC content, off-target locations and identities, and CHOPCHOP v2 activity scores (Table 4.1). RGEN-RFLP analyses revealed *in vitro* Cas9 cleavage efficiency for 11 of the *NAGLU* and *GBA1*-targeting gRNAs. Ten of 11 gRNAs direct Cas9 cleavage *in vitro* with >70% efficiency (Figures 4.2 – .5). Genomic sites of interest include *NAGLU* exon 2 (c.457 G>A), *NAGLU* exon 6 (c.1073 C>T), and *GBA1* exon 9 (c.1226 A>G) where three of four, three of three and four of four designed gRNAs showed >70% cleavage efficiency, respectively (Figures 4.2 – .5).

RGEN-RFLP analysis revealed the highest cleavage efficiency in *NAGLU* exon 2 was gRNAX with 88.1%, followed by gRNAY at 85.1%, then gRNAC with 70.4%. Minimal Cas9 cleavage efficiency was observed with gRNAU at 32.4%. The highest cleavage efficiency in *NAGLU* exon 6 was gRNAA with 95.2%, followed by gRNAG with 90.7% and gRNAB with 88.4%. *GBA1* gRNAs targeting the 55 bp deletion in exon 9 yielded 92% Cas9 cleavage efficiency for gRNA+49, followed by 88.7%, 88.4% and 83.0% for gRNA+58, gRNA+33, and gRNA+70, respectively.

Table 4.1 gRNA parameters for CRISPR-Cas9 HDR guides. sgRNAs targeting *NAGLU* c.889 C>T, c.1073 C>T, and c.457 G>A and *GBA1* c.1226 A>G. Mutation frequency and predicted disease severity are listed, as well as the proximity from the Cas9 cut site to the intended target base is indicated. Doench efficiency scores are determined through CHOPCHOP v2 (Labun *et al.*, 2019; Montague *et al.*, 2014; Doench *et al.*, 2016). Off-targets are also listed.

sgRNA	**Gene; Mutation**	**Mutation Frequency**	**Predicted disease severity (homozygous)**	**Cut site proximity to target base**	**GC Content (%)**	**Off-targets (1, 2, 3 MM)**	**Efficiency Score (CHOPCHOP; Doench *et al.*, 2016)**
gRNA1	*NAGLU*; c.889 C>T	12.5% (Weber *et al.*, 1999)	Severe (Christensen *et al.*, 2017)	-17	45	0, 0, 5	45.83
gRNA2				-36	50	0, 0, 4	48.17
gRNA3				+2	45	0, 0, 4	53.92
gRNAA	*NAGLU*; c.1073 C>T	Single report (Schmidtechen *et al.*, 1998)	Severe (Schmidtchen *et al.*, 1998; Clark *et al.*, 2019)	-8	65	0, 2, 10	55.93
gRNAB				0	80	0, 0, 8	43.92
gRNAG				-33	70	0, 0, 23	31.84
gRNAY	*NAGLU*; c.457 G>A	Single report (Schmidtechen *et al.*, 1998)	Severe (Schmidtchen *et al.*, 1998; Clark *et al.*, 2019)	-7	60	0, 0, 10	44.80
gRNAX				+29	70	N/A	N/A
gRNAU				-74	75	0, 0, 23	40.83
gRNAC				-31	60	0, 0, 7	53.80
gRNA+33	*GBA1*; c.1226 A>G	69.77% Jewish patients, 22.86% non-Jewish patients (Horowitz *et al.*, 1993)	Mild (Balwani *et al.*, 2011)	-33	55	0, 0, 6	59.59
gRNA+49				-49	65	0, 0, 6	18.61
gRNA+58				-58	55	0, 0, 2	22.77
gRNA+70				-70	45	0, 0, 2	43.23

T7EI assays revealed bands of expected sizes for three of three *NAGLU* c.889 C>T-targeting gRNA, gRNA1, -2, and -3, showed 2.6 ± 1.5%, 5.0 ± 1.1% and 2.0 ± 1.4% editing, respectively (Figure 4.2). Two of four *NAGLU* c.457 A>G gRNAs, gRNAY and gRNAC, show bands of expected sizes upon T7EI digest, with 20.6 ± 2.8% and 6.4% editing, respectively (Figure 4.3). gRNAY was further characterized for the presence of indels via *in silico* ICE analysis, revealing 27% indels (Figure 4.3). gRNAs targeting *NAGLU* exon 6 and *GBA1* exon 9 were unable to be confirmed via T7E1 assays, despite multiple attempts to transfect patient-derived iPSCs with RNP complexes containing these gRNAs. Based on these results, gRNA2 and gRNAY were chosen for knock-in experiments with CRISPR HDR.

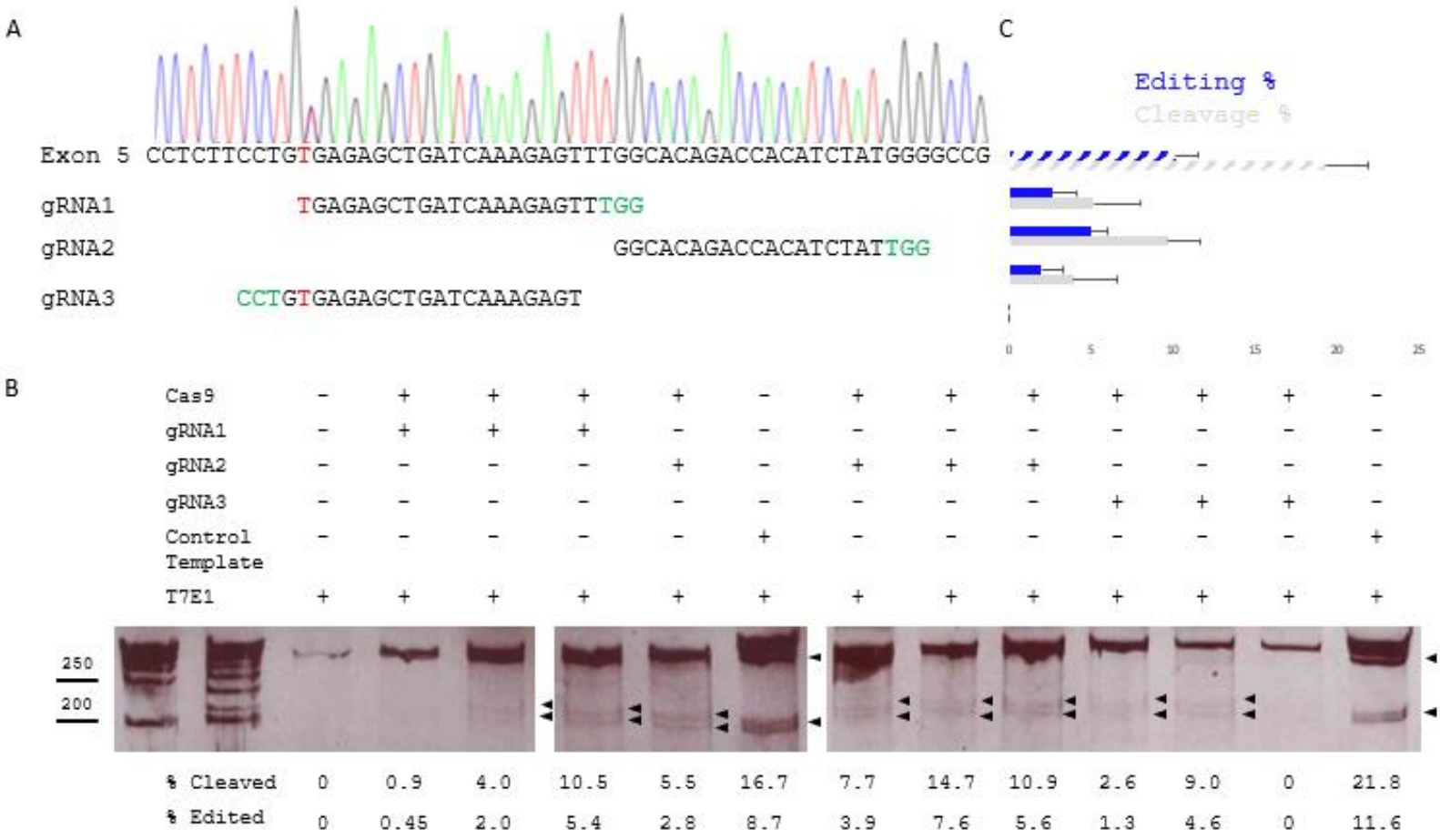
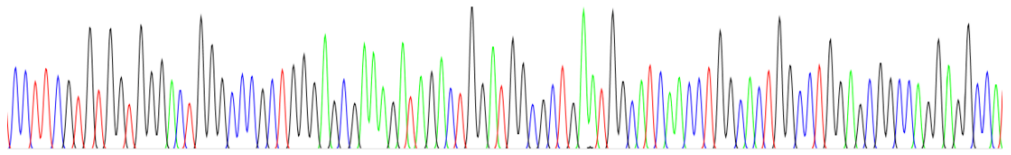


Figure 4.2 gRNA testing for c.889 C>T *NAGLU* target. Three gRNAs (gRNA1, -2, -3) were designed to target the c.889 C>T mutation in exon 5 of *NAGLU*. **A.** DNA chromatogram showing heterozygous c.889 C>T mutation in exon 5. gRNA sequences are listed below aligned with the *NAGLU* exon 5 target site. **B.** These gRNAs were delivered to patient-derived iPS cells using lipofection and lysed 48 hours later. T7EI assays were performed on *NAGLU* exon 5 PCR products amplified from the cell lysates and run on 8% polyacrylamide gels. T7EI recognizes heteroduplexed DNA, indicative of indels. Cleaved products range from 177 – 257 bp and are indicated with black arrows. **C.** Cleavage and editing efficiencies are represented by grey and blue bars for gRNA1, -2, and -3, with 2.6 ± 1.5% (n = 3), 5.0 ± 1.1% (n = 4) and 2.0 ± 1.4% (n = 3; SEs represented) editing, respectively. Dashed bars represent positive control cleavage and editing efficiencies for T7E1 activity. gRNA2 was determined to have the highest editing efficiency (5.0 ± 1.1%) and was used in HDR experiments going forward.

A



Exon 2 CCTTCGTGTGGTGGGACTGGGCCCGCTGGGAGCGAAAGATAGACTGGATGGCGCTGAATGGCATCAACCTGGCACTGGCCTGGCACTGGCCTGGAGCGGCCAGGAGGCCA

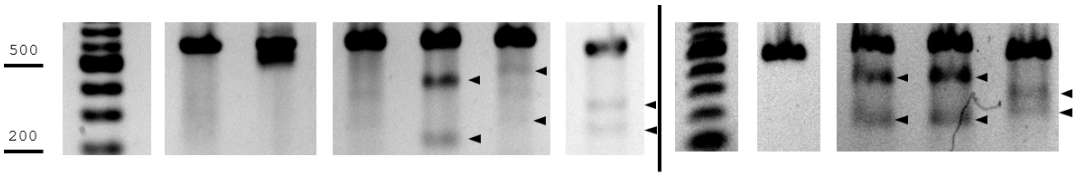
gRNAY GACCCTCGCTTTCTATCTGATGG

gRNAX CCTTCGTGTGGTGGGACTGGGCC

gRNAU GGAGCGGCCAGGAGGCCATCAGG

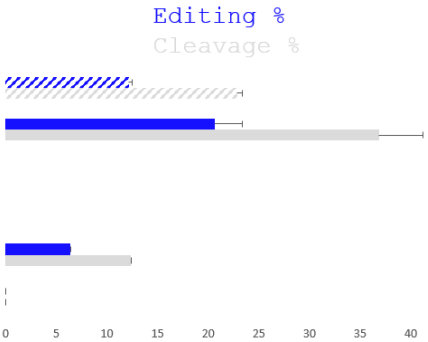
gRNAC GGCGCTGAATGGCATCAACCTGG

Cas9	+	+	+	+	+	–	+	+	+	–	
gRNAX	+	+	–	–	–	–	–	–	–	–	
gRNAU	–	–	+	–	–	–	–	–	–	–	
gRNAY	–	–	–	+	–	–	+	+	+	–	
gRNAC	–	–	–	–	+	–	–	–	–	–	
Control Template	–	–	–	–	–	+	–	–	–	+	
T7E1	–	+	+	+	+	+	–	+	+	+	

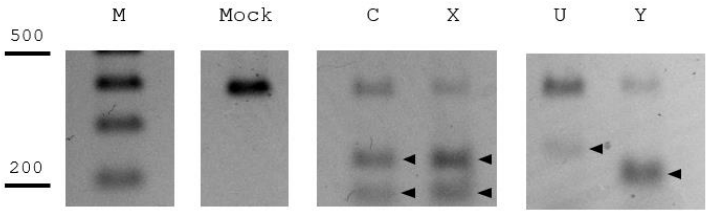


% Cleavage	0	N/A	N/A	29.8	12.4	22.3	0	44.8	36.0	23.4
% Editing	0	N/A	N/A	16.2	6.4	11.9	0	25.7	20.0	12.5

C



B



% Cleavage	0	70.4	88.1	32.4	85.1
% GC	0	60	70	75	55

D

Figure 4.3 gRNA testing for c.457 G>A *NAGLU* target. Four gRNAs (gRNAY, -C, -U, and -X) were designed to target the c.457 G>A mutation in exon 2 of *NAGLU*. **A.** DNA chromatogram showing homozygous c.457 G>A mutation in exon 2. gRNA sequences are listed below aligned with the *NAGLU* exon 2 target site. These gRNAs were delivered to patient-derived iPSCs using lipofection and lysed 48 hours later. T7EI assays were performed on *NAGLU* exon 2 PCR products amplified from the cell lysates and run on 8% polyacrylamide gels. T7EI recognizes heteroduplexed DNA, indicative of indels. Cleaved products range from 159 – 436 bp and are indicated with black arrows. **B.** RGEN-RFLP analysis reveals Cas9 *in vitro* cleavage efficiency is the highest for gRNAX with 88.1%, followed by gRNAY with 85.1%, then gRNAC with 70.4%. Minimal Cas9 cleavage efficiency was observed with gRNAU at 32.4%. Cleaved products are indicated with black arrows. **C.** Cleavage and editing efficiencies are represented by grey and blue bars for gRNAY, -X, -U, and -C, with 20.6 ± 2.8% (n = 3), 0% (n = 1), 6.4% (n = 1) and 0% (n = 1; SEs represented) editing, respectively. Dashed bars represent positive control cleavage and editing efficiencies for T7E1 activity. gRNAY was determined to have the highest editing efficiency via T7EI (20.6 ± 2.8%). Synthego ICE analysis revealed 27% indels upon gRNAY-directed Cas9 cleavage. gRNAYwas used in HDR experiments going forward(**D.**).

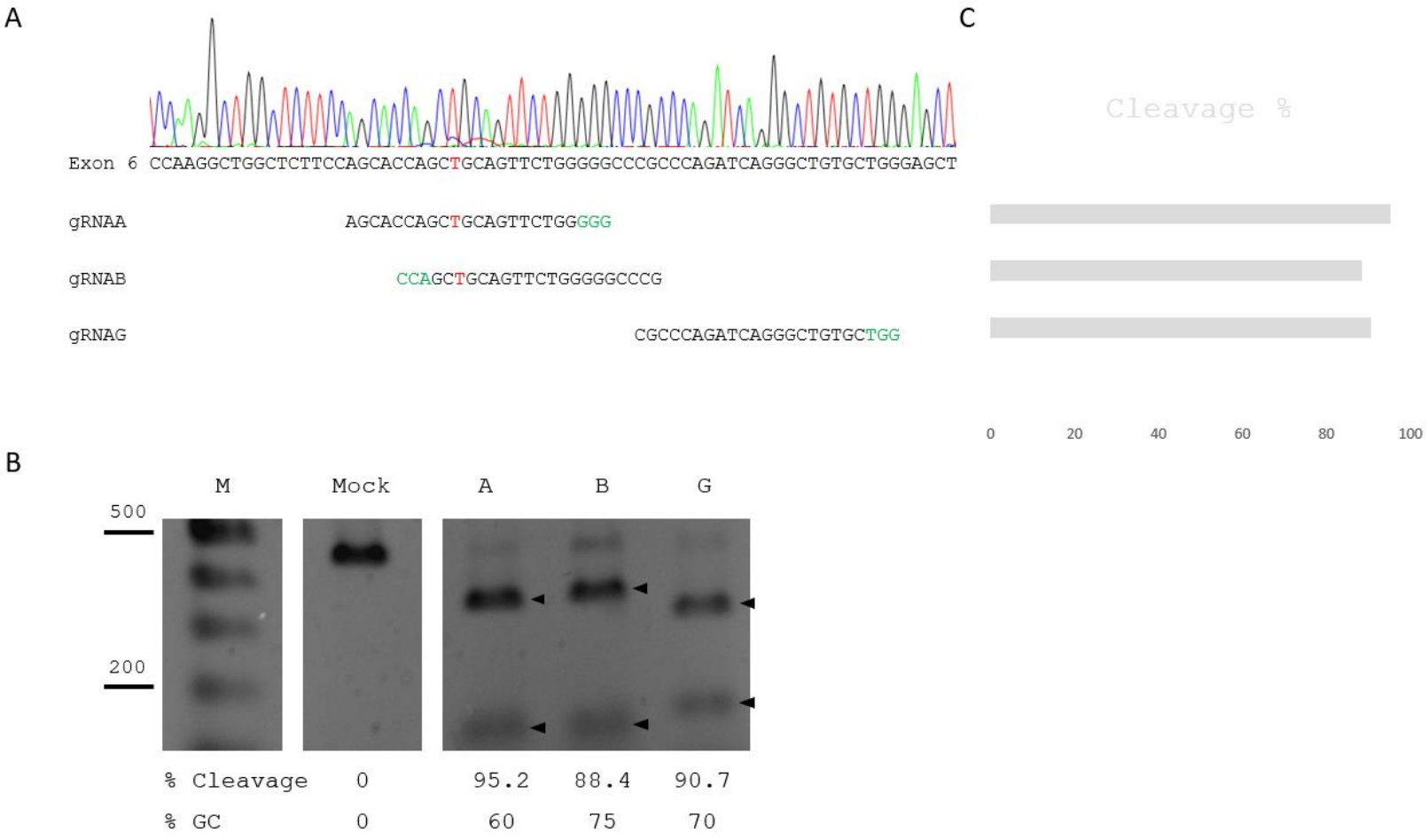


Figure 4.4 gRNA testing for c.1073 C>T *NAGLU* target. Three gRNAs (gRNAA, -B, and -G) were designed to target the c.1073 C>T mutation in exon 6 of *NAGLU*. **A.** DNA chromatogram showing homozygous c.1073 C>T mutation in exon 2. gRNA sequences are listed below aligned with the *NAGLU* exon 6 target site. **B.** RGEN-RFLP analysis reveals Cas9 *in vitro* cleavage efficiency is the highest for gRNAA with 95.2%, followed by gRNAG at 90.7%, then gRNAB with 88.4% and cleavage products are indicated by black arrows. **C.** Cas9 cleavage efficiencies are represented by grey bars.

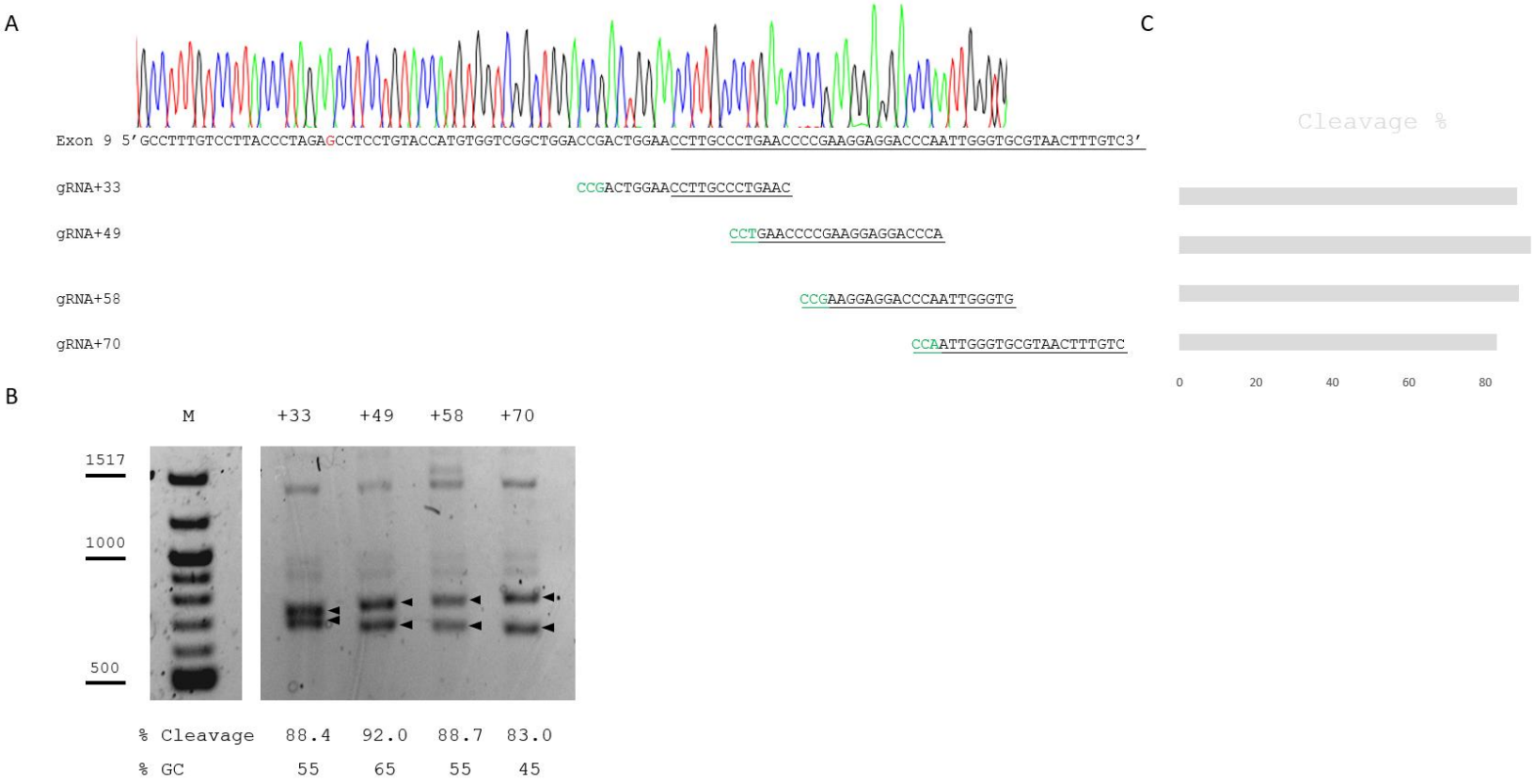


Figure 4.5 gRNA testing for c.1226 A>G *GBA1* target. Four gRNAs (gRNA+33, +49, +58, and +70) were designed to target the c.1226 A>G mutation in exon 9 of *GBA1*. **A.** DNA chromatogram showing homozygous wildtype c.1226A in exon 2*. gRNA sequences are listed below aligned with the *GBA1* exon 9 55 bp deletion target site. **B.** RGEN-RFLP analysis reveals Cas9 *in vitro* cleavage efficiency is the highest for gRNA+49 with 92.0%, followed by gRNA+58 at 88.7% and gRNA+33 at 88.4, then gRNA+70 with 83.0% and cleavage products are indicated by black arrows. **C.** Cas9 cleavage efficiencies are represented by grey bars. *Chromatogram showing sequence extending into the region of the 55 bp deletion is unavailable for c.1226 A>G homozygous genotype. The c.1226 A>G homozygous mutation was confirmed via Sanger sequencing of a shorter amplicon that does not encompass the target sites of the listed gRNAs here.

4.3.1.2. Base editing design constructs

A gRNA was designed to target the *GBA1* mutation c.1226 A>G. Recommendations for successful guide design include a GC content between 40 and 60%, and a transition mutation of interest 13-17 nucleotides upstream from a suitable 5'-NGG-3' PAM sequence (Figure 4.6). However, gRNA with GC contents outside of this range can also result in successful on-targeting. A gRNA targeting *HPRT1*, utilized here with the base editing system, was previously validated to have an on-target efficiency of 10.4% using CRISPR-Cas9, measured using T7EI analysis (Sakuma *et al.*, 2014).

Table 4.2. sgRNA for base editing in *GBA1* and *HPRT1*. Target gene, mutation prevalence, predicted disease severity, and GC content of the sgRNA sequence are indicated. Additionally, the target base for deamination within the base editing window relative to the PAM site is indicated.

sgRNA	Gene; Mutation	Mutation Prevalence	Predicted disease severity (homozygous)	Application	Target base position in window	GC Content (%)
BE_*GBA1*	*GBA1*; c.1226 A>G	69.77% Jewish patients, 22.86% non-Jewish patients (Horowitz *et al.*, 1993)	Mild (Balwani *et al.*, 2011)	Base editing	+5	50
BE_*HPRT1*	*HPRT1*	N/A	N/A		+1, +2, +3, +5	30

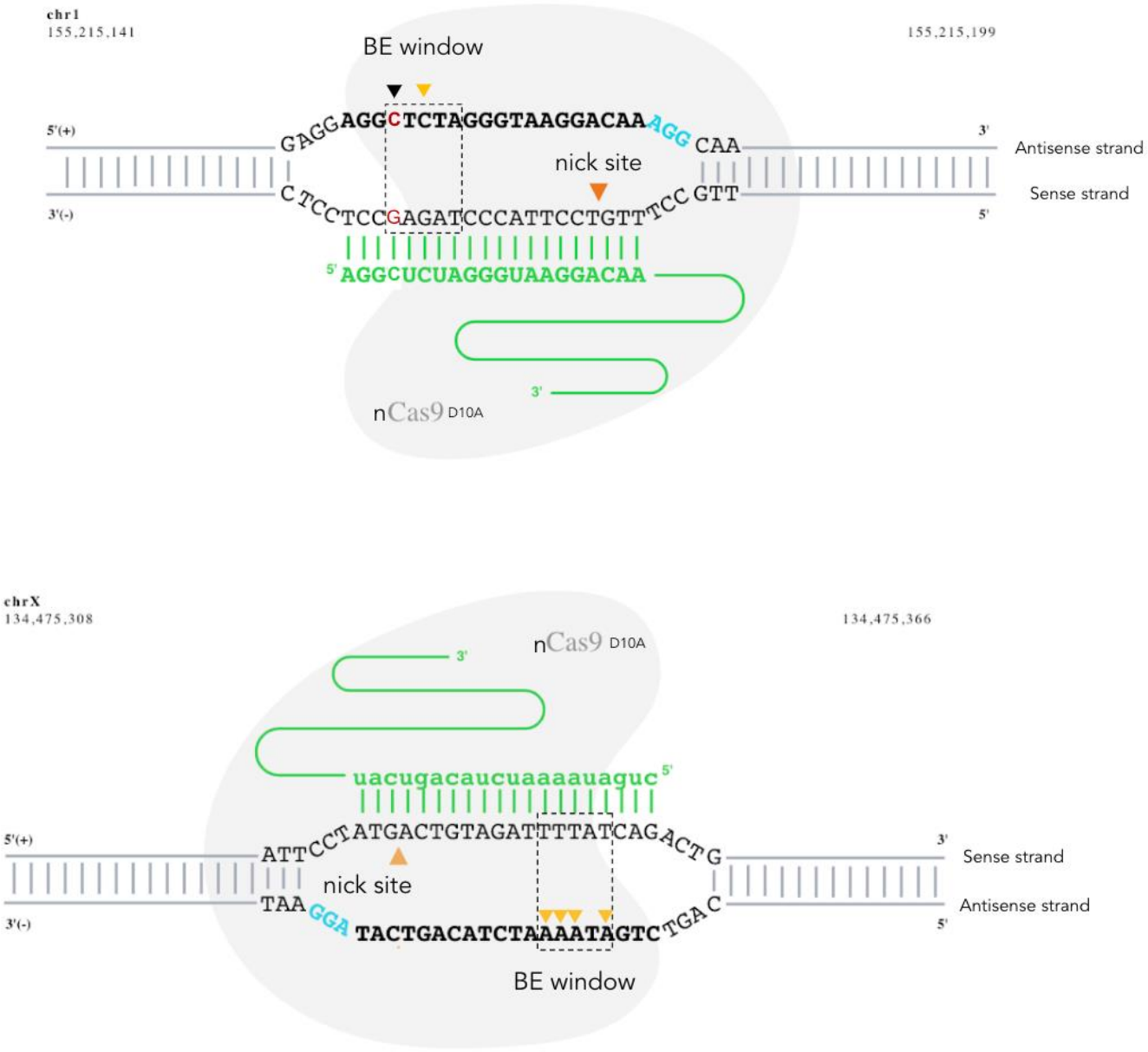


Figure 4.6. Base editor designs for *GBA1* c.1226 A>G mutation (top), and *HPRT1* control (bottom). nCas9 (grey) is directed to a site of interest using a gRNA (green) where a nick is created (orange arrows) three nucleotides upstream of a PAM site (blue). An ABE or CBE (not pictured) will deaminate target bases (black arrows) within the base editing window (dashed box). Bystander bases are indicated by yellow arrows. Base editor schematics are modified from Synthego CRISPR Design Tool.

4.3.1.3. Prime editing design constructs

A pegRNA for use with the prime editing system was designed using the *in silico* pegFinder design tool (Chow *et al.*, 2020), following design specifications outlined in Anzalone *et al.* (2019). This pegRNA targets the *GBA1* mutation c.1226 A>G. A positive control pegRNA (Addgene plasmid #132778) that installs a CTT insertion in *HEK3* was also obtained.

Table 4.3. pegRNA target mutation information for prime editing in *GBA1*. Target gene, mutation prevalence, predicted disease severity, and GC content of the spacer component of the pegRNA sequence are indicated. Additionally, the target mutation relative to the PAM site is indicated.

pegRNA	Gene; Mutation	Mutation Prevalence	Predicted disease severity (homozygous)	Application	Mutation location relative to PAM	GC Content (%)
c1226AS-5	*GBA1*; c.1226 A>G	69.77% Jewish patients, 22.86% non-Jewish patients (Horowitz *et al.*, 1993)	Mild (Balwani *et al.*, 2011)	Prime editing	-5	60

Table 4.4 pegRNA design for prime editing in *GBA1*. The underlined portion of the 3' extender is the wildtype base to be incorporated in place of the mutated base. The italicized nucleotide is the PAM site silent mutation.

pegRNA	GC (%)	pegRNA spacer forward oligo (5' → 3')	pegRNA spacer reverse oligo (5' → 3')	3' extender and primer binding site forward oligo (5' → 3')	3' extender and primer binding site reverse oligo (5' → 3')	Invariable scaffold forward oligo (5' → 3')	Invariable scaffold reverse oligo (5' → 3')
c1226AS-5	60	cacc*g*AGCCGACCAC ATGGTACAGGgtttt	ctctaaaacCCTGTA CCATGTGGTCGGCTc	gtgcACCCTAGA<u>A</u>*<u>T</u>*C TCCTGTACCATGTGGT	aaaaACCACATGGTA CAGGAGA<u>T</u>TCTAGGGT	AGAGCTAGAAATAGC AAGTTAAAATAAGGC TAGTCCGTTATCAACT TGAAAAAGTGGCACC GAGTCG	GCACCGACTCGGTGC CACTTTTTCAAGTTGA TAACGGACTAGCCTTA TTTTAACTTGCTATTTC TAG

4.3.2. Plasmid construction and bacterial transformation

gRNA and pegRNA sequences were ligated into scaffolding plasmids (Addgene plasmids #43860 and #132777). Plasmids were found to have the correct ligated oligonucleotide sequence as confirmed by Sanger sequencing (Supplementary Figures 4.1 – .3).

4.3.3. Mammalian cell transfections

4.3.3.1. Lipofection

Lipofection was optimized using *HPRT1* editing efficiencies with a range of RNP concentrations and lipofectamine volumes. The optimal lipofection condition was determined to be 30 nM with 1.5 μL lipofectamine yielding a 15.9% editing efficiency (Figure 4.7). Successful lipofection was confirmed via ATTO550™ fluorescence (Figure 4.8).

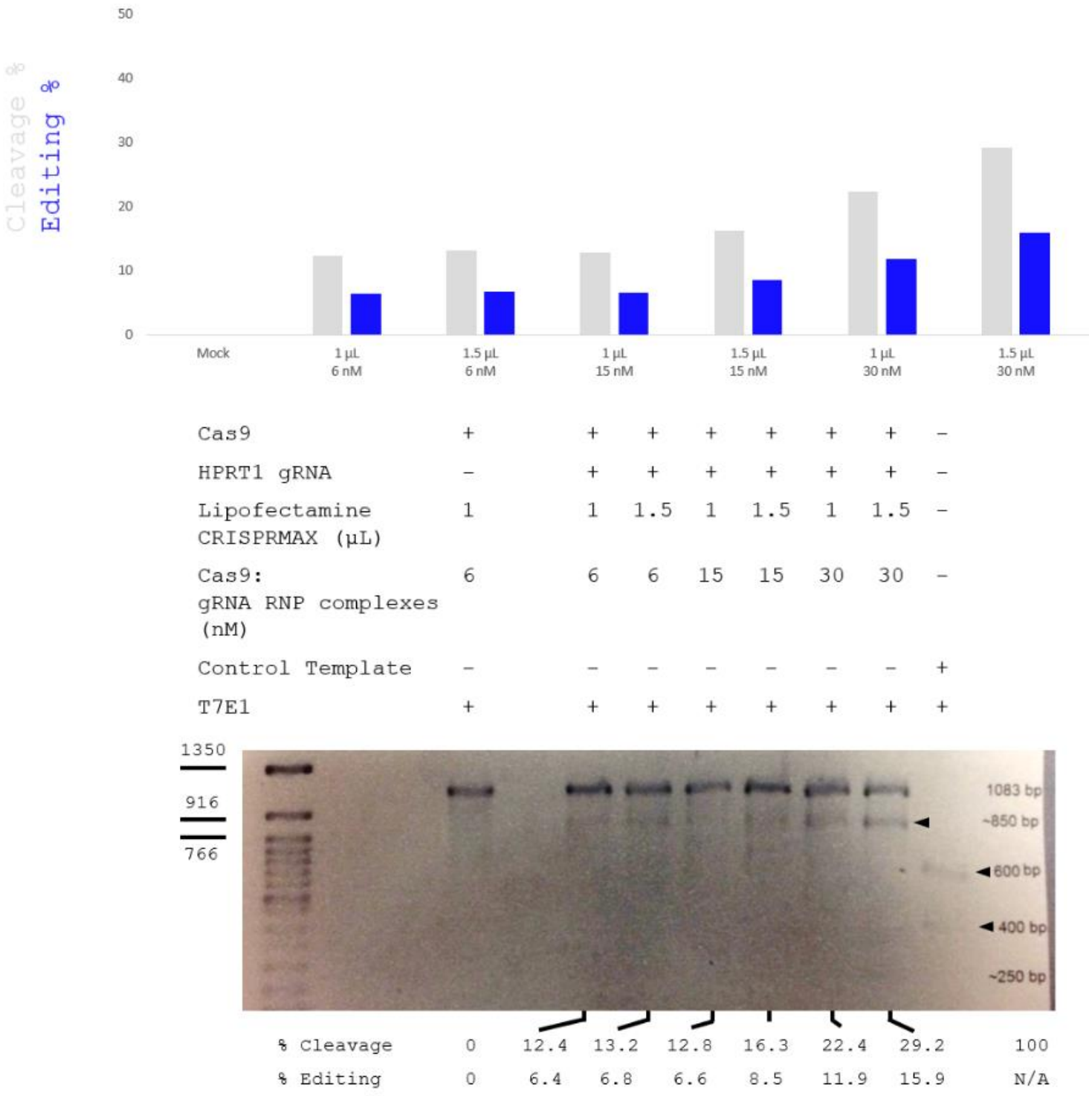


Figure 4.7. Optimization of lipofection conditions in human iPSCs. gDNA was isolated from homozygous c.457 G>A iPSCs lipofected with RNPs targeting the *HPRT1* gene. Amplified DNA was subjected to a T7EI assay and run on a 2% agarose gel. Densitometry was used in

ImageJ v2 (Rasband, 2008) to determine editing and cleavage efficiencies by comparing digest products to the uncleaved band. Densitometry of cleaved amplicons revealed the highest editing efficiency of 15.9% when 30 nM RNPs and 1.5 μL lipofectamine is used.

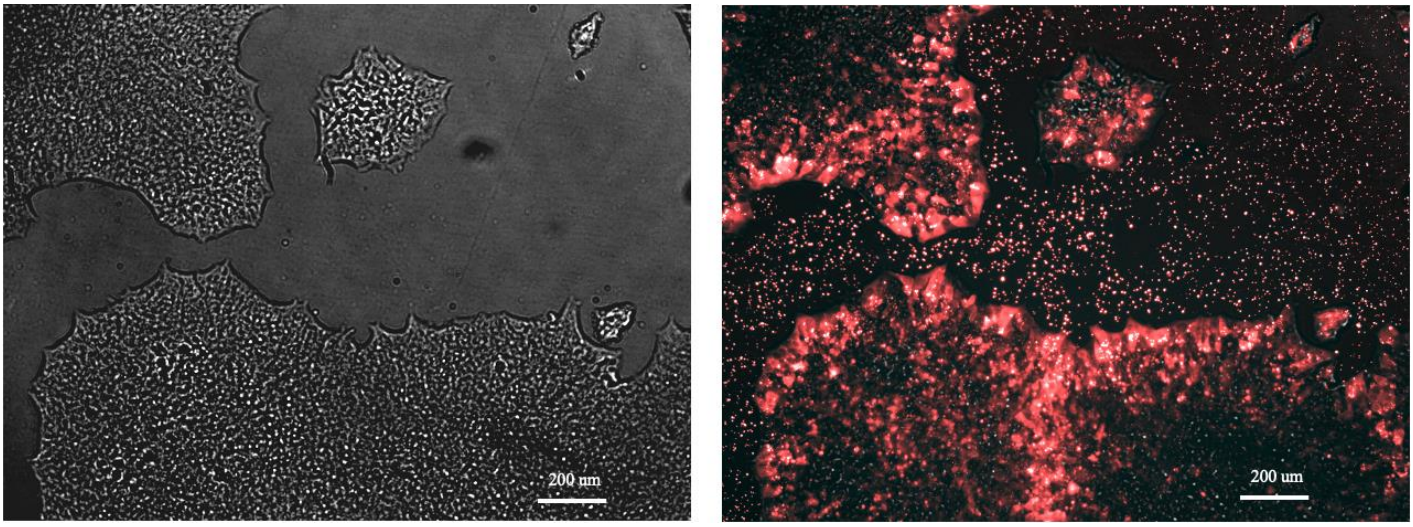


Figure 4.8. Fluorescence and bright field photomicrographs of human iPSCs 24 hours post-lipofection. ATTO550™ complexed to gRNAs is used as a marker of transfection success. Fluorescence and brightfield photomicrographs were obtained using a Leica DMI3000 B inverted microscope and QCapture™ Pro 7 Software. Fluorescence micrographs were colourized from greyscale in ImageJ v2 (Rasband, 2008) and overlayed with brighfield photomicrographs (left) yielding a composite image (right).

4.3.3.2. Electroporation

Electroporation was optimized for iPSCs using four different protocols previously reported for stem cells. Of the four tested conditions, 1200 V, 20 ms, and 2 pulses yielded the highest transfection efficiency (19.3 ± 6.6%) and cell viability (33.6 ± 6.5%) across three replicates. This electroporation protocol was used going forward (Table 4.5, Figure 4.9)

Table 4.5. A comparison of electroporation conditions on normal human iPSCs. A comparison of electroporation parameters on normal human episomally reprogrammed iPSCs. Test conditions were chosen based on reported electroporation protocols (Thermo Fisher Scientific) resulting in high cell viability and transfection efficiency in cell types similar to iPSCs. Protocol 1 (1400 V/ 10 ms/ 3 pulses), 2 (1200 V/ 30 ms/ 1 pulse), 3 (1100 V /30 ms/ 2 pulses), and 4 (1200 V/ 20 ms/ 2 pulses) were found to result in cell viabilities and transfection efficiencies of 20.9 ± 1.2% and 14.7 ± 14%, 27.0 ± 3.2% and 13.4 ± 6.8%, 7.3 ± 2.7% and 11.3 ± 8.5%, and 33.6 ± 6.5% and 19.3 ± 6.6%, respectively (n = 3). A plasmid encoding YFP was used to assess transfection efficiency through fluorescence imaging. Images were analysed using ImageJ v2 (Rasband, 2008) to determine total transfection efficiency for each condition.

Protocol No.	Protocol	Cell Type	Cell viability (%)	Transfection efficiency (%)
1	1400 V/ 10ms/ 3 pulse	Adipose-derived SCs	96	88
		Human iPSCs	**20.9 ± 1.2**	**14.7 ± 14**
2	1200 V/ 30 ms/ 1 pulse	BGO1V Human ESCs	92	33
		Human iPSCs	**27.0 ± 3.2**	**13.4 ± 6.8**
3	1100 V/ 30ms/ 2 pulses	H9 human ESCs	99	29
		Human iPSCs	**7.3 ± 2.7**	**11.3 ± 8.5**
4	1200 V/ 20ms/ 2 pulses	ADSCs, BGO1V human ESCs, mouse ESCs	99, 90, 99	79, 21,79
		Human iPSCs	**33.6 ± 6.5**	**19.3 ± 6.6**

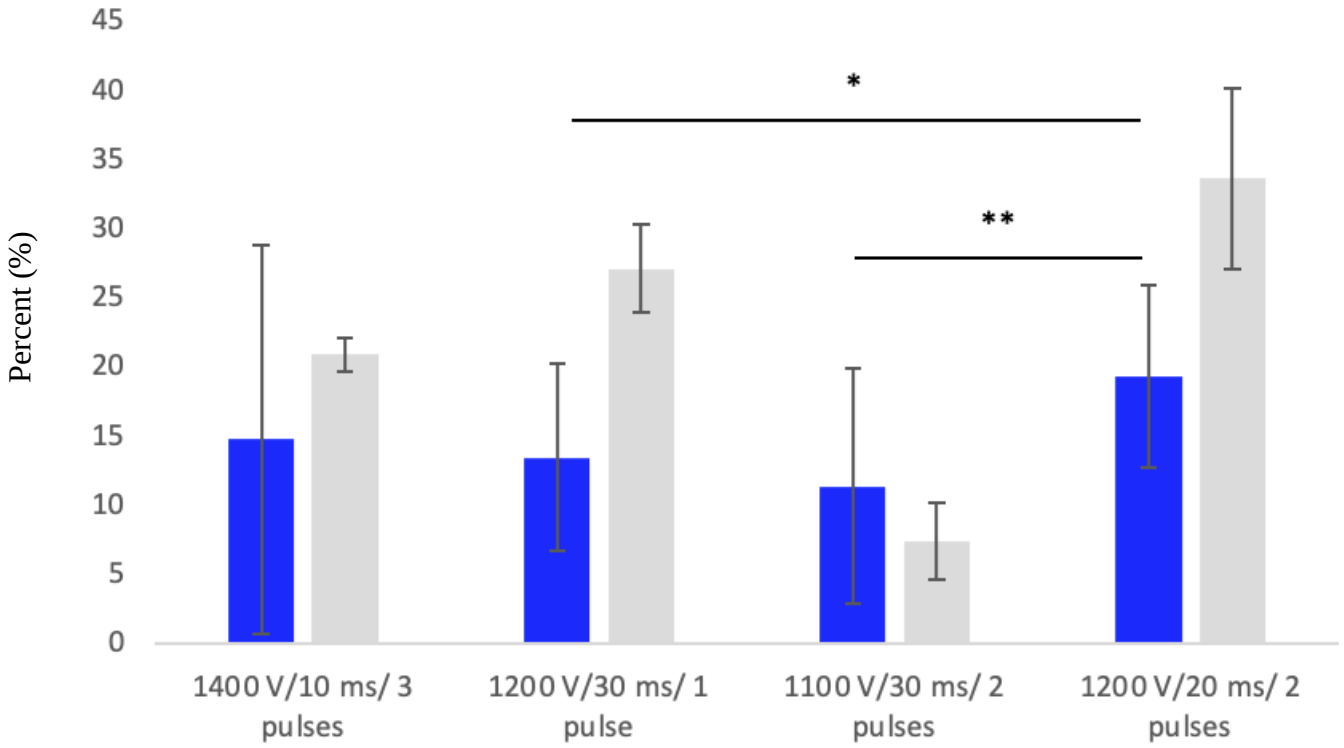


Figure 4.9. A comparison of electroporation conditions on normal human iPSCs. A comparison of electroporation parameters on normal human episomally reprogrammed iPSCs. Test conditions were chosen based on reported electroporation protocols (Thermo Fisher Scientific) resulting in high cell viability and transfection efficiency in cell types similar to iPSCs. Protocol 1 (1400 V/ 10 ms/ 3 pulses), 2 (1200 V/ 30 ms/ 1 pulse), 3 (1100 V /30 ms/ 2 pulses), and 4 (1200 V/ 20 ms/ 2 pulses) were found to result in cell viabilities and transfection efficiencies of 20.9 ± 1.2% and 14.7 ± 14%, 27.0 ± 3.2% and 13.4 ± 6.8%, 7.3 ± 2.7% and 11.3 ± 8.5%, and 33.6 ± 6.5% and 19.3 ± 6.6%, respectively (n = 3). A plasmid encoding YFP was used to assess transfection efficiency through fluorescence imaging. Images were analysed using ImageJ v2 (Rasband, 2008) to determine total transfection efficiency for each condition. Comparisons not shown for electroporation efficiency are not significant. No statistical comparisons are shown for cell confluency. Cells were seeded at 4.3 x 10^5 cells/well, where a confluent well is 2.4 x 10^5 cells/well. Cell viability = cell confluence/1.79. Cell viability is represented by grey columns, transfection efficiency is represented by blue columns. P values were calculated using student's t-test with two tails and heteroscedasticity. Assumption of normality was made, but q-q plots were not generated or analysed. *p val <0.05, **p val <0.01.

4.3.4. Genome editing outcomes and *in silico* analyses

4.3.4.1. CRISPR-Cas9

A mutagenized forward primer was designed to amplify exon 5 of *NAGLU* while introducing an adenine in place of a cytosine in order to create a restriction site in the presence of the c.889 C>T mutant sequence (see chapter 2, Figure 2.3). iPSC clones were isolated between 2 and 30 days after lipofection with gRNA2, ssODN2 and Cas9, and propagated on 96 well plates. Of the first 48 colonies isolated, 41 survived the monoclonal selection (85.4%). Of 362 isolated clones, 72 (19.9%) were chosen for Sanger sequencing. This decision was made based on the densitometry being at or above four standard deviations of the mean density ratio of the undigested to digested band in control conditions (66.1 ± 2.3%, n = 13). Therefore, only clones that exceeded 75.3% undigested band intensity were sent for Sanger sequencing. Of these 72 clones, none showed knock-in of c.889 C in *NAGLU* exon 5. Similarly, c.457 G>A isolated monoclones yielded no detectable HDR. Using a similar method with a *Bco*DI restriction enzyme (discussed in chapter 2), c.457 G>A editing was also analyzed, yielding no detectable HDR upon digests and Sanger sequencing.

4.3.4.2. Base editing

Base editing outcomes in *HPRT1*

For base editing of HPRT1, a gRNA that creates a base editing window with four adenines was selected. Overall editing efficiency at this locus was determined to be 6 ± 0.5% (n = 3, Figure 4.10). Review of the Sanger sequencing DNA chromatogram data suggested that the majority of editing occurred at the center of the base editing window. Editing efficiency was then analyzed for each position within the base editing window. Across three trials, position +1 and +2 show no editing, position +3 yielded 27 ± 4.3% editing, and position +5 yielded 3 ± 1.5% (n = 3, Figure 4.10). HEK293 cells transfected with an ABE plasmid and gRNA plasmid targeting *HPRT1* show a transfection efficiency of 51.7% ± 15.5% (Figure 4.10).

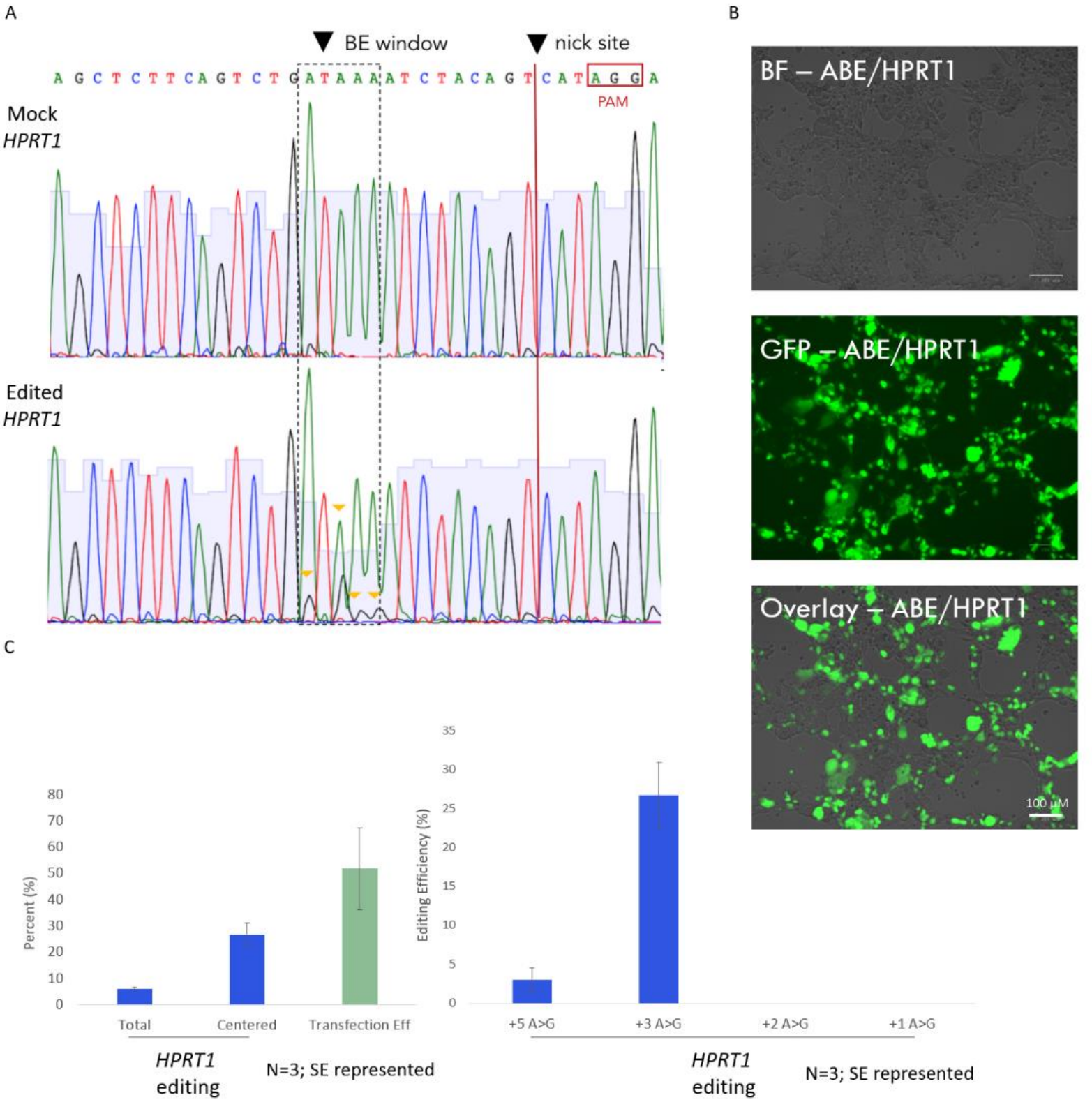

A
BE window
nick site
A G C T C T T C A G T C T G A T A A A A T C T A C A G T C A T A G G A
PAM
Mock
HPRT1
Edited
HPRT1
B
BF – ABE/HPRT1
GFP – ABE/HPRT1
Overlay – ABE/HPRT1
100 µM
C
Percent (%)
80
70
60
50
40
30
20
10
0
Total
Centered
Transfection Eff
HPRT1
editing
N=3; SE represented
Editing Efficiency (%)
35
30
25
20
15
10
5
0
+5 A>G
+3 A>G
+2 A>G
+1 A>G
HPRT1
editing
N=3; SE represented

Figure 4.10 Base editing outcomes in *HPRT1* in HEK293 cells. A. DNA chromatograms show WT (Mock) *HPRT1* sequence on chromosome X. The base editing gRNA directs Cas9 nicking (red line) three bases upstream from the PAM sequence (5'-NGG-3') and the base editing window (dashed lines) is 13-17 nucleotides from the PAM sequence. Target bases are marked with yellow arrows. **B.** Fluorescence and brightfield photomicrographs obtained using a ZOE™ Fluorescence Microscope (Bio-Rad) show a transfection efficiency of 51.7 ± 15.5%. **C.** *HPRT1* editing and transfection efficiencies represented graphically (left), and *HPRT1* editing efficiencies at specific target nucleotide location within the base editing window (right). Measured from the PAM site, position +1 and +2 show no editing, whereas position +3 and +5 show 27% and 3% editing efficiency, respectively (n = 3). Editing efficiencies were determined using ICE analysis. Scale bars represent 100 μm.

Base editing outcomes in *GBA1*

GD patient iPSCs and skin FBs, as well as HEK293 cells, were targeted using a CBE plasmid (pCMV_BE4max_P2A_GFP, Addgene #112099) and the abovementioned p.N370S-targeting gRNA with an editing efficiency of 4%, 42% and 54%, respectively, when cultured post electroporation for 96 hours (Figure 4.12). HEK293 cells cultured for an additional 24 hours showed 95% base editing (Figure 4.12). Transfection efficiency was calculated across HEK293, c.1226 A>G FBs, and c.1226 A>G iPSCs and determined to be 48.7 ± 8.5%, 38.0 ± 4.9%, and 3.4 ± 0.8%, respectively (Figure 4.11).

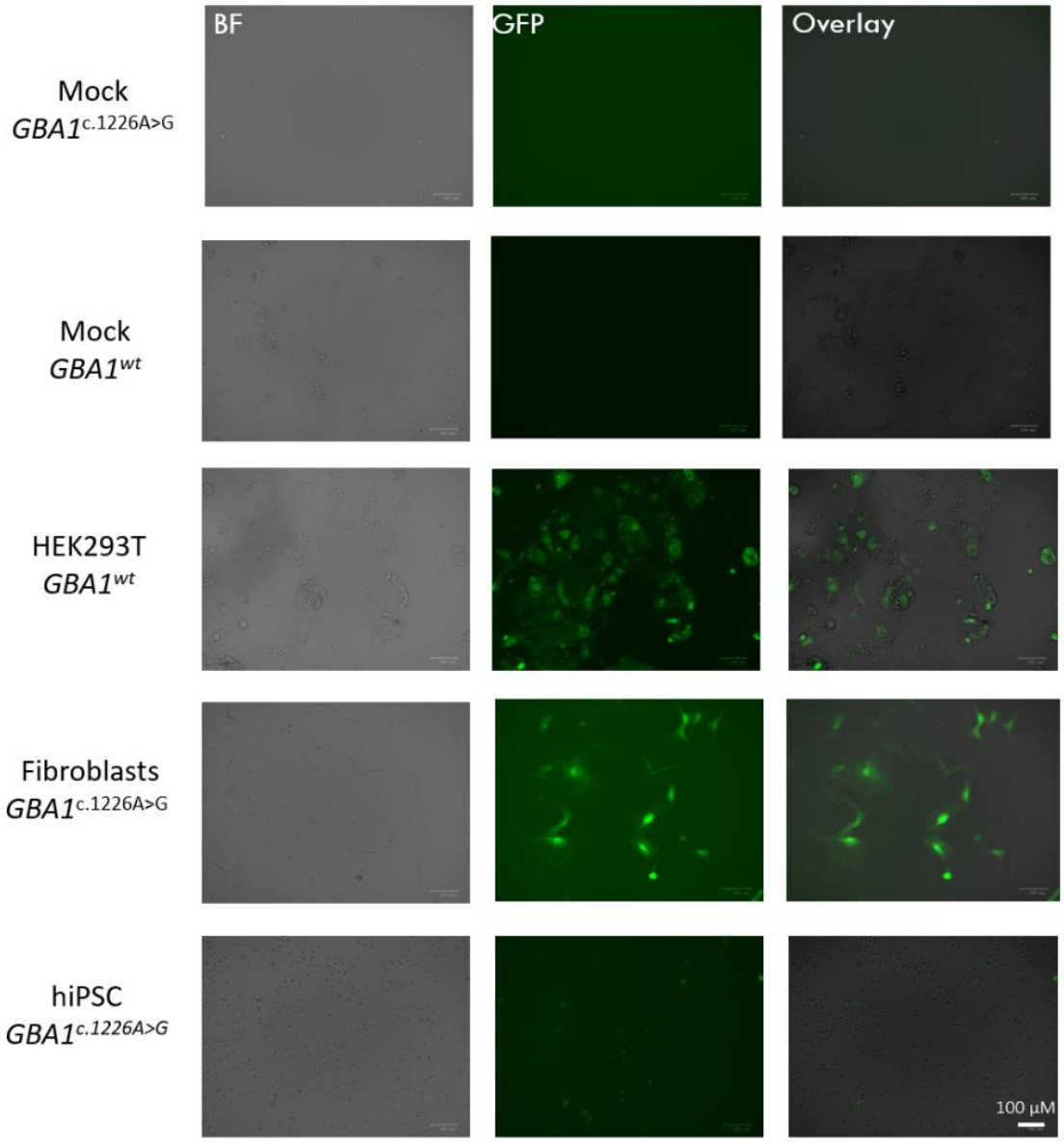


Figure 4.11. Transfection efficiencies in HEK293, c.1226 A>G FBs, and c.1226 A>G iPSCs. Brightfield and fluorescence photomicrographs depicting GFP transfection marker expression across three cell types (left) were obtained using a ZOE™ Fluorescence Microscope (Bio-Rad). Cells were electroporated at 1200 V, 20 ms, 2 pulses to deliver gRNA and CBE-containing plasmids. Transfection efficiencies were determined to be 48.7 ± 8.5%, 38.0 ± 4.9%, and 3.4 ± 0.8%, respectively (n = 1). Standard errors are calculated based on 10 photomicrographs analysed per condition. Scale bars represent 100 μm.

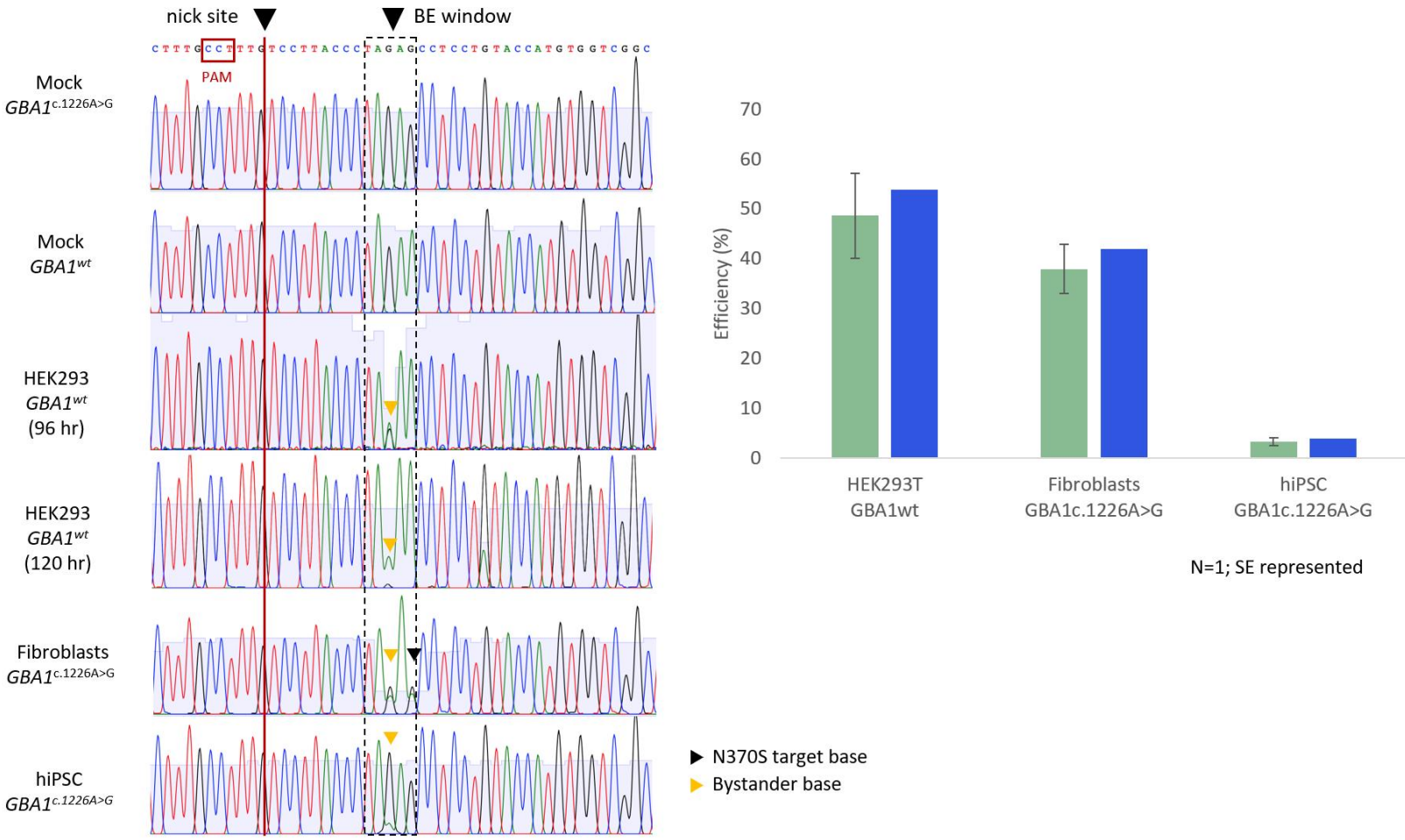


Figure 4.12. Base editing outcomes in *GBA1* in HEK293, c.1226 A>G FBs, and c.1226 A>G iPSCs. DNA chromatograms show mock wildtype and c.1226 A>G mutant sequence on chromosome 1q21.1 *GBA1* exon 9. The antisense base editing gRNA creates a nick (red line) three bases upstream from the PAM sequence (5'-NGG-3') and the base editing window (dashed lines) is 13-17 nucleotides from the PAM sequence. Bystander bases are marked with yellow arrows, and target bases are marked with black arrows. Editing and transfection efficiencies are compared (right), where blue columns represent editing efficiency and green columns represent transfection efficiency. For HEK293, c.1226 A>G FBs, and c.1226 A>G iPSCs editing efficiency is 54%, 42%, and 4%, respectively. When culture time post-electroporation was extended to 120 hours, 95% editing efficiency was obtained in HEK293 cells. For HEK293, c.1226 A>G FBs, and c.1226 A>G iPSCs transfection efficiency is 48.7 ± 8.5%, 38.0 ± 4.9%, and 3.4 ± 0.8%, respectively. Editing efficiencies were determined using ICE analysis. Editing efficiencies are calculated for the bystander base only in HEK293 cells, and for the c.1226 G>A editing efficiency only in p.N370S-containing alleles. Standard errors are calculated.

4.3.4.3. Prime editing

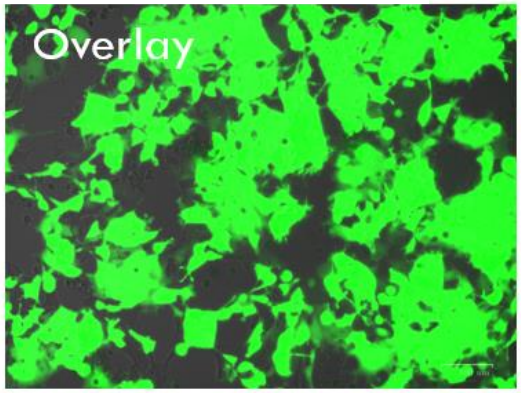
Prime editing was conducted in HEK293, FBs, and iPSCs using a CTT insertion pegRNA. CTT was installed with an installation efficiency of 11.8%, 0% and 5.1%, respectively, when cultured for 96 hours post-transfection (Figure 4.13). CTT installation efficiency was determined for HEK293 cells every 24 hours for a period of four days. No editing was observed at 24 or 48 hours (data not shown). Editing efficiency was determined to be 10.8 ± 0.2% and 12.5 ± 0.5% for 72 and 96 hours, respectively (n = 2; Figure 4.14).

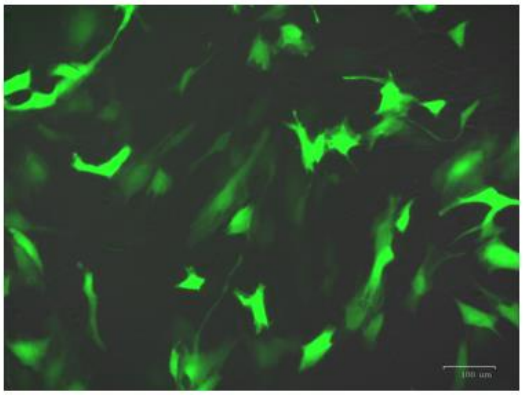
A

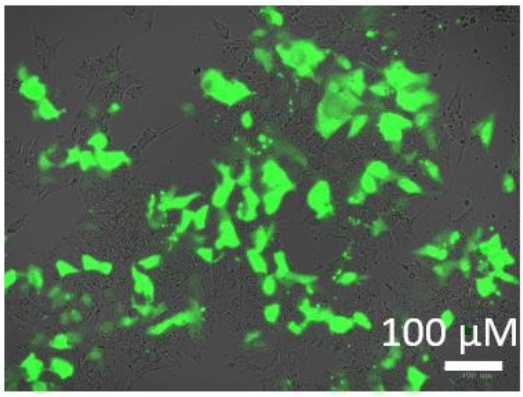
HEK 293

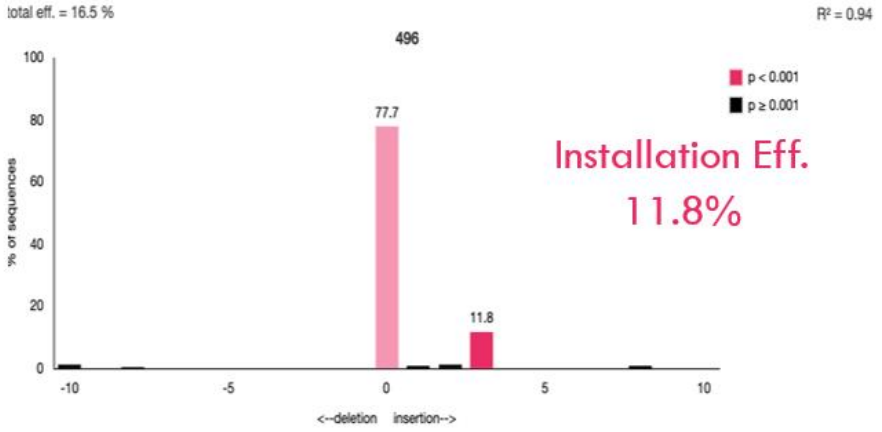
FB

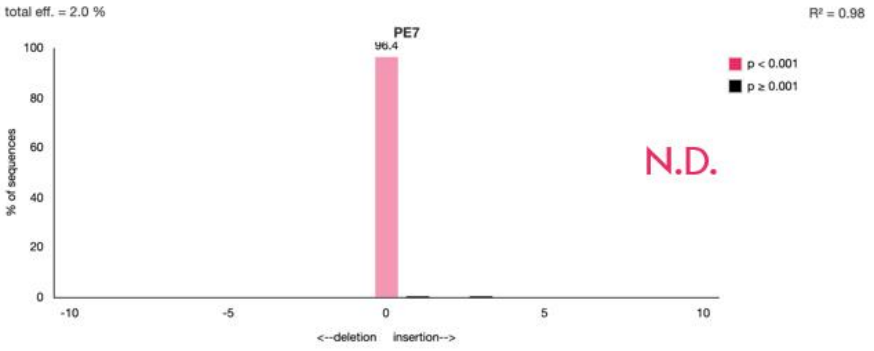
iPS

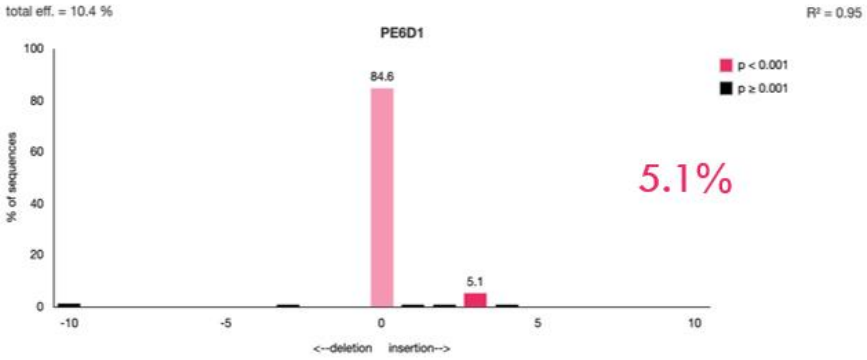












B

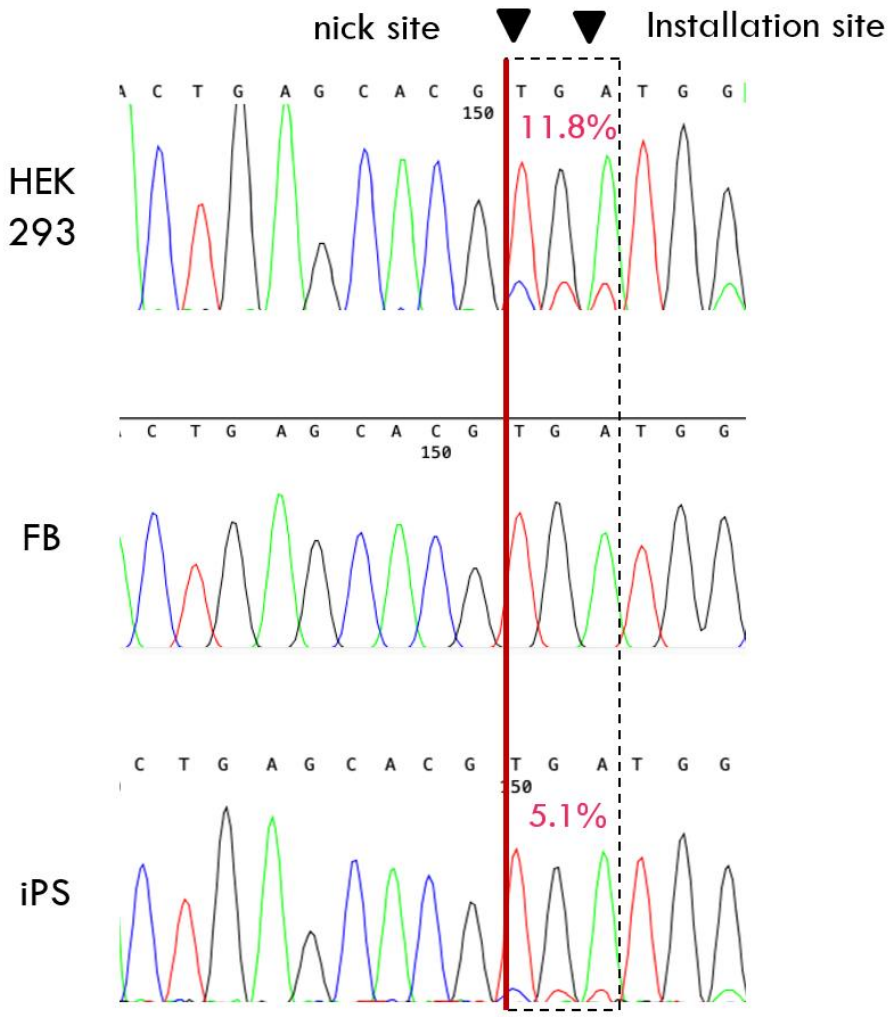


Figure 4.13. Installation of CTT edit in *HEK3* using the PE2 system. A. Top, middle and bottom panels show HEK293 cells, FBs, and iPSCs, respectively, transfected with pCMV-Venus positive control plasmid, a proxy for transfection efficiency. Replicate wells contained PE2-encoding and CTT installation pegRNA plasmids (Anzalone *et al.*, 2019). Percent CTT installation was determined using TIDE analysis, results shown as bar graphs (middle column). DNA chromatograms are shown in (**B.**) CTT is installed immediately downstream from the Cas9 nick site. Scale bars represent 100 μm.

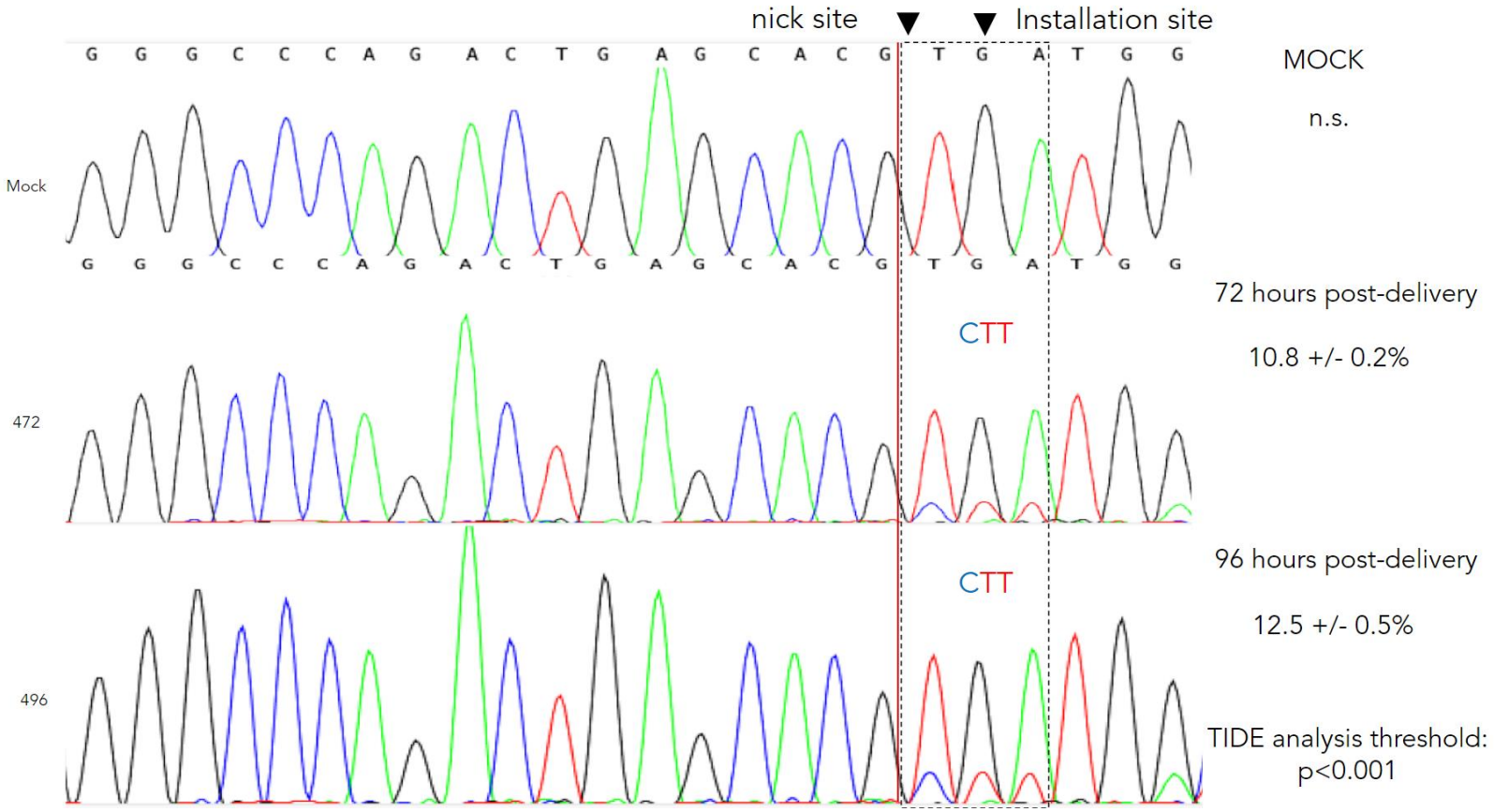


Figure 4.14. DNA chromatograms showing CTT installation in *HEK3* over time. HEK293 cells were electroporated with plasmids encoding PE2 and a CTT installation pegRNA (Anzalone *et al.*, 2019), then cultured to time points 24 (not shown), 48 (not shown), 72 and 96 hours. CTT installation efficiency was undetectable for 24 and 28 hours, and 10.8 ± 0.2% and 12.5 ± 0.5% for 72 and 96 hours, respectively, post-transfection, respectively. CTT is installed immediately downstream from the nCas9 nick site. CTT installation efficiencies were determined using *in silico* TIDE analysis.

Experimental conditions were performed alongside YFP-positive transfection control condition. Transfection efficiencies were determined by measuring fluorescence using ImageJ v2

(Rasband, 2008) at 24 hours post-electroporation. HEK293, FBs, and iPSCs were transfected with an efficiency of 82.3% ± 2.9%, 22.5% ± 2.4% and 38.7 ± 6.3%, respectively (Figure 4.15 – .16).

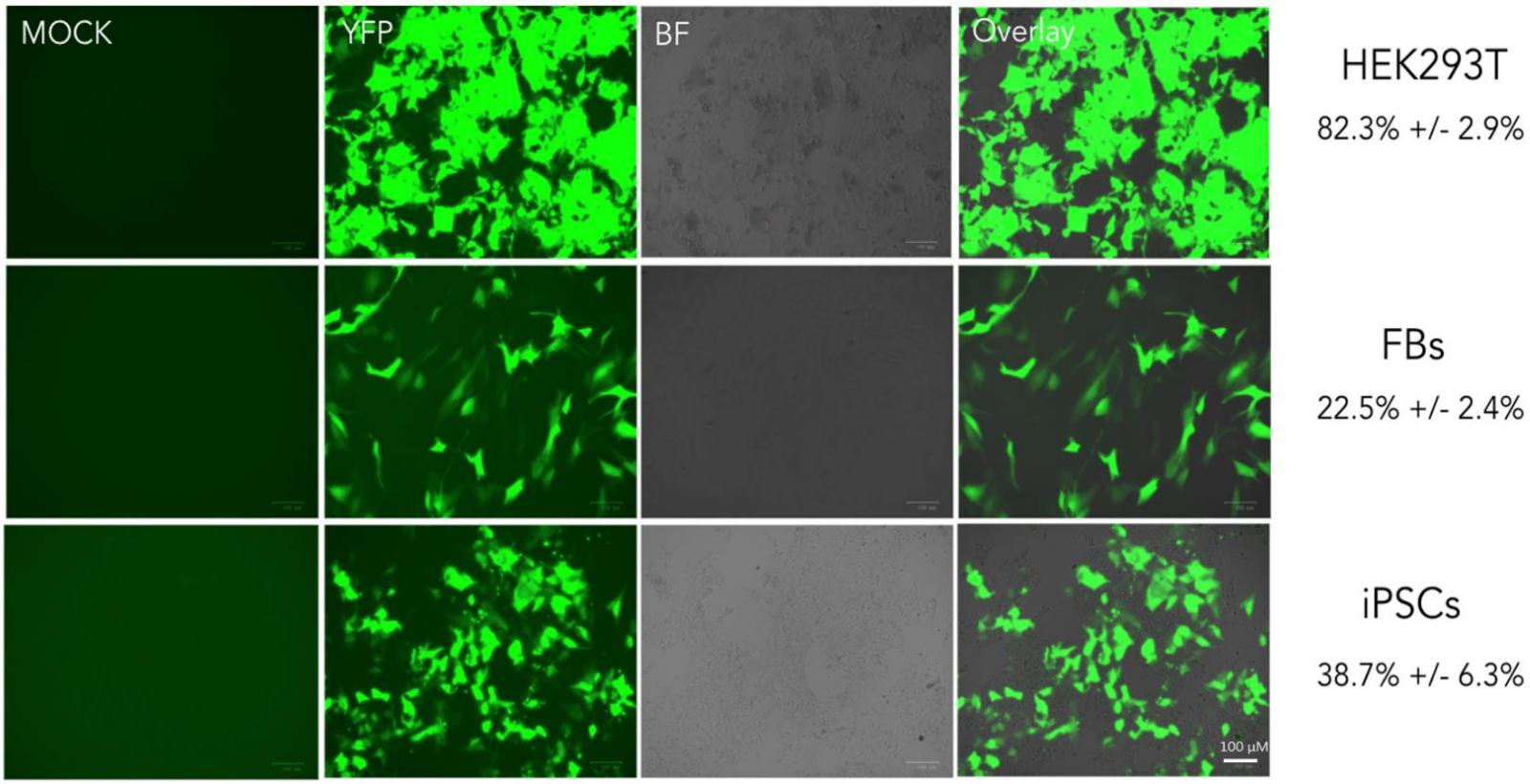


Figure 4.15 Transfection efficiency photomicrographs for HEK293, FBs, and iPSCs. Cells were transfected with a YFP-containing plasmid. Images were taken with a ZOE™ fluorescence microscope (Bio Rad) at 24 hours post-electroporation. HEK293, FBs, and iPSCs were transfected with an efficiency of 82.3% ± 2.9%, 22.5% ± 2.4% and 38.7 ± 6.3%, respectively. Transfection efficiencies were calculated using ImageJ v2 (Rasband, 2008) by calculating total fluorescence on 10 separate fluorescence photomicrographs and compared to total cell confluence on counterpart brightfield photomicrographs. Scale bars represent 100 µm.

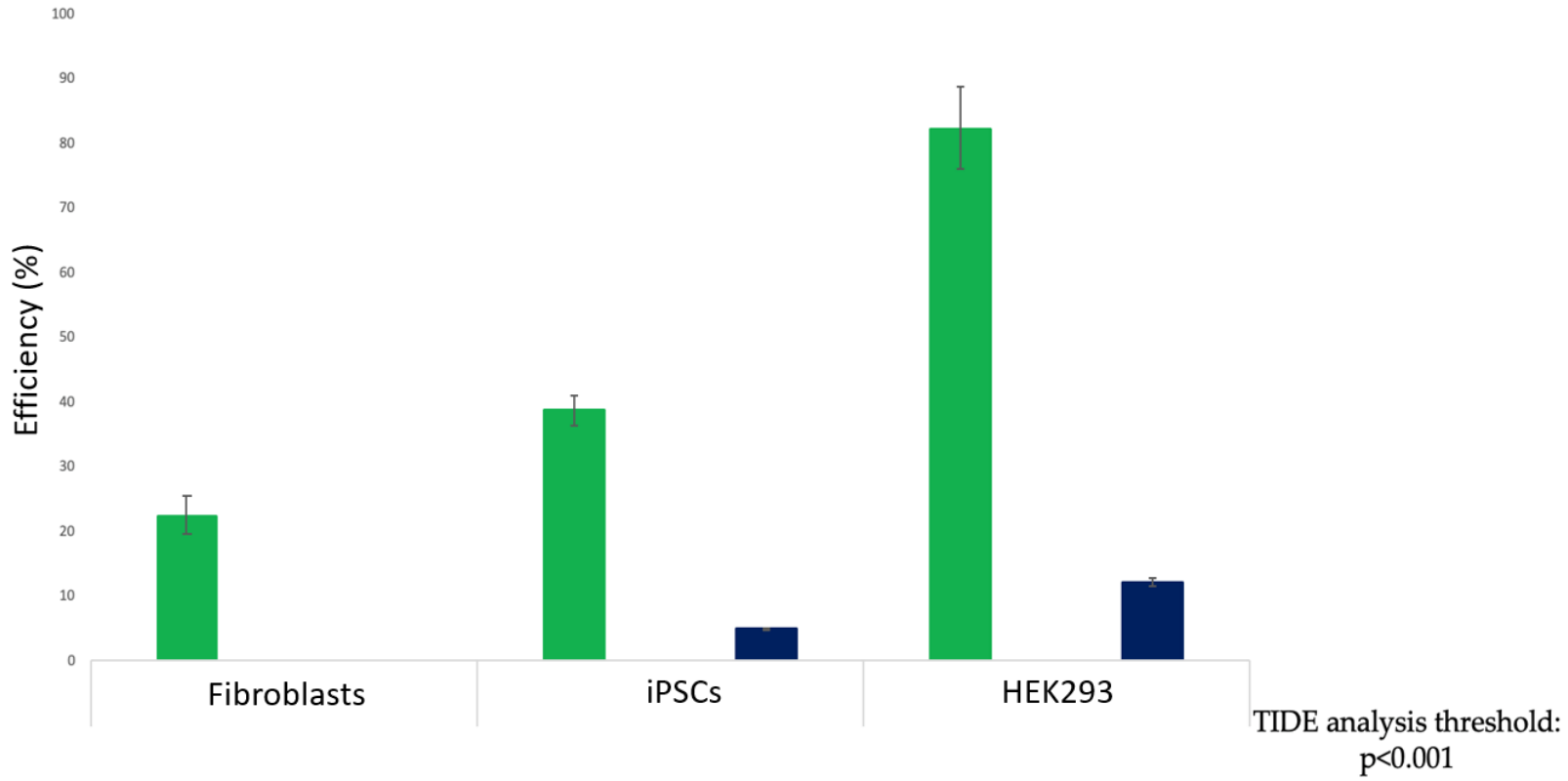


Figure 4.16 A comparison of transfection efficiency and prime editing installation efficiency across three cell types: FBs, iPSCs, and HEK293 cells. FBs, iPSCs, and HEK293 cells were transfected with an efficiency of 22.5% ± 2.4%, 38.7 ± 6.3%, and 82.3 ± 2.9%, respectively. CTT was installed in FBs, iPSCs, and HEK293 cells with an efficiency 0%, 5.1%, and 11.8%, respectively. Transfection efficiencies were calculated using ImageJ v2 (Rasband, 2008) by calculating total fluorescence on 10 separate fluorescence photomicrographs and compared to total cell confluence on counterpart brightfield photomicrographs. Green bars represent transfection efficiency and blue bars represent installation efficiency. Installation efficiencies were calculated using TIDE analysis, $p<0.001$.

Prime editing in *GBA1*

Preliminary prime editing at *GBA1* targeting the c.1226 A>G mutation shows 1% PAM site alteration after *in silico* analysis with Synthego ICE (Figure 4.17).

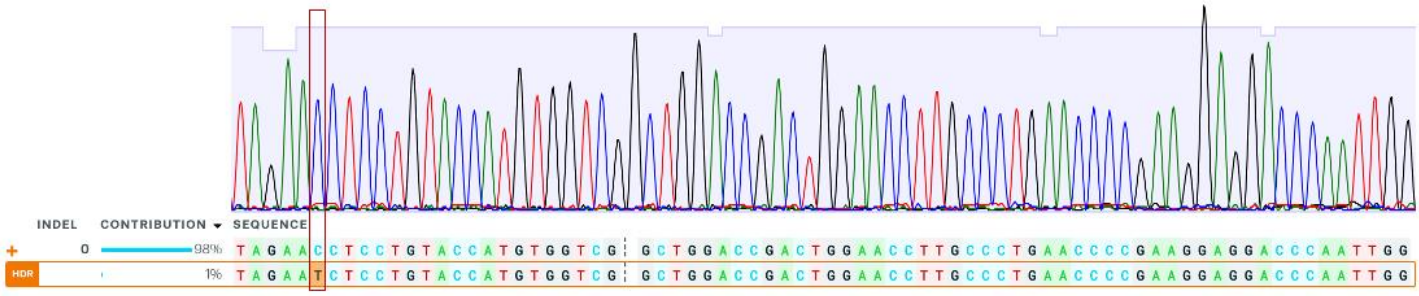


Figure 4.17. Preliminary PAM silent mutation installation in HEK293 cells using prime editors and pegRNA targeting the c.1226 A>G mutation. HEK293 cells were transfected with PE2-encoding and c.1226 A>G pegRNA plasmids. Transfected HEK293 cells were lysed after 96 hours, and *GBA1* was amplified using previously described *GBA1*-specific PCR protocols. *In silico* analyses conducted via Synthego ICE. PAM site silent mutation installation was detected at a 1% editing efficiency. Location of PAM site silent mutation is indicated by a red box.

4.4. Discussion

We sought to design gRNAs and pegRNAs to target disease-causing mutations in *NAGLU* and *GBA1*, and test these designs *in vitro* with CRISPR HDR, base editing, and prime editing systems. The goal was to establish genome editing strategies capable of reverting patient-specific point mutations causative for MPS IIIB and GD to their normal genotype. Edited patient-specific mutations *in vitro* are predicted to recapitulate molecular and biochemical characteristics within the edited cells that mimic the functions of healthy cells. These proof-of-concept techniques will set the stage for creating *in vitro* disease models and may translate to *in vivo* systems.

Here, we describe outcomes of the CRISPR-Cas9 system to target mutations in *NAGLU* (c.457 G>A, c.889 C>T, and c.1073 C>T) and *GBA1* (c.1226 A>G) *in vitro*. gRNAs were designed with a number of considerations in mind: (1) Cas9 cut site proximity to the point mutation of interest; (2) Doench on-target activity score of >0.2 (Doench *et al.*, 2016); and (3) ability to silently mutate the PAM sequence to prevent Cas9 re-targeting. A number of gene-specific considerations were also imperative. *NAGLU* has a high overall GC content, particularly in the second and sixth exons (Schmidtchen *et al.*, 1998; Selmer *et al.*, 2012). This increases the number of putative PAM (5'NGG3') sequences available, although gRNAs with very low or high GC content are not as efficient at targeting DNA, with 40-70% is the recommended range (Labun *et al.*, 2016; Labun *et al.*, 2019; Montague *et al.*, 2014). An additional complication in the case of *GBA1* targeting is the presence of a pseudogene, *GBAP1*, 16 kb downstream from *GBA1* that shares 96% sequence similarity with *GBA1* (Martinez-Arias *et al.*, 2001). In order to avoid large chromosomal deletions and decreased efficiency while targeting c.1226 A>G, gRNAs can be designed to avoid targeting *GBAP1* and to instead target a 55 bp region of DNA proximal to c.1226 that is deleted in the pseudogene (Christensen and Choy, 2017). This requirement may reduce the efficiency of genome editing in *GBA1* since the closest possible sgRNA targeting c.1226 A>G has a Cas9 cut site 33 nucleotides downstream from the c.1226 A>G mutation.

Two experimental approaches were used to analyze on-target efficiency of gRNAs for use with CRISPR-Cas9. The first approach was RNA-guided endonuclease RFLP (RGEN-mediated RFLP) analysis (Kim *et al.*, 2014). This method was originally proposed to improve the identification of indels and mutations found in some cancers (Kim *et al.*, 2014). This approach

requires an RNP and functions as a reverse RFLP: wildtype DNA is recognized but mutant DNA is not. Kim *et al.* demonstrate this technique for mutation detection and as an alternative to mismatch detection assays, such as T7EI and Surveyor nucleases, as a means for genotyping, as well as identifying RGEN-mediated editing (2014). We chose to adapt the RGEN-mediated RFLP technique to aid in the confirmation and validation of gRNA design. Here, we show that RGEN-mediated RFLP can be utilized as a rapid, cursory analysis of gRNA targeting *in vitro* prior to target cell transfection. We found that Cas9 cleavage efficiency for 10/11 gRNAs is directed *in vitro* with >70% efficiency. Interestingly, gRNAX showed a dramatically different targeting efficiency between the RGEN-RFLP (88.1%) and T7EI assays (no detectable indels). This may be due to the high GC content of gRNAX (70%) and the fact that this is an unranked gRNA in CHOPCHOP, suggesting that this gRNA was not predicted to be capable of driving appreciable Cas9 cleavage in a cellular context (Table 4.1)

One of the 11 gRNAs tested resulted in only 32.4% cleavage efficiency after a RGEN-RFLP (gRNAU), and indeed resulting in no detectable indels upon transfection and T7EI assay. This is the first description of Cas9 RNPs being used as a direct RGEN-RFLP prior to testing gRNAs within the cellular context. Similar to PCR primers, gRNA function is dependent on parameters such as lack of self-complementarity (Hsu *et al.*, 2014), secondary structure (Liu *et al.*, 2016), favourable GC content (Bruegmann *et al.*, 2019), and repetitive nucleotides (Wong *et al.*, 2015; Wang *et al.*, 2019). As more sophisticated algorithms are designed for predicting gRNA on-targeting efficiency, further necessary design parameters will be elucidated. For example, a preference for a thymine immediately 3’ of the PAM has been noted to result in a higher Cas9 activity (Doench *et al.*, 2016). However, individual gRNAs with high predicted Cas9 on-target

activity have been reported to have low efficiency *in vitro* (Huang *et al.*, 2020). Therefore, it is important to test Cas9 activity with multiple gRNAs prior to proceeding to HDR experiments.

Next, T7E1 assays were performed to determine CRISPR Cas9 on-target efficiency within the context of mammalian cells, in this case iPSCs. This mismatch detection assay utilizes the bacteriophage resolvase T7EI, which is capable of recognizing >2 bp indels and cleaving at these sites (Babon *et al.*, 2003; Vouillot *et al.*, 2015). In preparation for T7EI digestion, PCR amplicons generated from the amplification of gDNA from a pool of edited cells are denatured and reannealed, resulting in heteroduplex formation (Babon *et al.*, 2003). Since not all indels are recognized and since not all denatured DNA strands will reanneal with a mismatched DNA strand, the T7EI assay underrepresents the total amount of genome editing that has occurred (Vouillot *et al.*, 2015). The highest on-targeting efficiency for *NAGLU* exon 5 was obtained with gRNA2 (5.0 ± 1.1%), which has a GC content of 50%, minimal off-targets; however, the Cas9 cut site is located 36 nucleotides away from the c.889 C>T point mutation in exon 5. gRNAY targeting of the c.457 G>A mutation in *NAGLU* exon 2 had the highest apparent on-target efficiency of 20.6 ± 2.8% as determined using the T7EI assay; however, ICE analysis captured an increased indel proportion of 27% of Sanger sequencing reads (Synthego Performance Analysis, ICE Analysis. 2019. V2.0. Synthego; [September 18th, 2020]). ICE produces NGS-quality results from Sanger sequencing data. NGS is far more accurate at determining the on-target efficiency of gRNAs than T7EI assays, for the former is not limited by indel size nor does NGS rely on stochastic denaturation and reannealing for accurate indel detection (Sentmanat *et al.*, 2018). gRNAY overlaps the target point mutation and the Cas9 cut site is located seven nucleotides upstream from the mutation.

Despite numerous attempts to transfect patient iPSCs with Cas9 and gRNAs targeting *GBA1* and *NAGLU* exon 6 and detect on-targeting via T7EI assays, on-targeting was unable to be attained from any of our gRNAs targeting c.1226 A>G in *GBA1* or c.1073 C>T in *NAGLU*. The six exons of *NAGLU* have a high GC content, making gRNA design particularly challenging (Schmidtchen *et al.*, 1998; Selmer *et al.*, 2012). gRNA designs that direct Cas9 cleavage within a reasonable distance from the c.1073 C>T mutation have GC contents of 65-80% (Table 4.1), close to or outside the upper range of suggested GC content for predicted efficient gRNA (Labun *et al.*, 2016; Labun *et al.*, 2019; Montague *et al.*, 2014). This higher GC content may explain the lack of on-targeting using these gRNAs. Based on the described RGEN-RFLP, T7EI, and ICE findings, gRNA2 and gRNAY were chosen for knock-in experiments with CRISPR HDR.

It is important to design and test a number of gRNAs for each genome editing strategy in order to elucidate guides that are not only predicted to target a site accurately and efficiently via current *in silico* algorithms, but also capable of directed Cas9 catalyzed DSBs within the cellular context. Closed chromatin greatly decreases gRNA binding and Cas9 cleavage (Daer *et al.*, 2017; Isaac *et al.*, 2016). It has been found that areas of genes present in nucleosomes are unable to be targeted by Cas9 (Horlbeck *et al.*, 2016).

To identify edited iPSC clones, transfected iPS monoclonal colonies were isolated 2-30 days post-transfection and propagated in duplicate on 96-well plates. Amplified gDNA was initially screened using RFLP. Exon 5 c.889 C>T putative edited clones were analyzed using *Hpy*CH4III in a reverse RFLP: putative editing was indicated by a lack of digestion. For exon 2, *Bco*DI was used in a RFLP to identify putative edited monoclones. Both methods utilized

mismatch PCR to create a restriction site. Clones with banding patterns that indicate possible HDR were sent for Sanger sequencing. CRISPR HDR was undetectable in *NAGLU* exon 5 and 2 upon Sanger sequencing. There are a few explanations as to why no editing may have been obtained upon CRISPR HDR. Human iPSCs are hard-to-transfect cells, resulting in increased cell death (Chatterjee *et al.*, 2011; Tan *et al.*, 2019) and reduced transfection efficiency (Tan *et al.*, 2019) in comparison to transformed cell lines (Maurisse *et al.*, 2010). Here, we utilized a lipofection transfection method for CRISPR HDR in iPSCs using a reverse transfection method, plating iPSCs and transfecting simultaneously (Yu *et al.*, 2016). Reverse transfection has reportedly increased overall transfection efficiency (Yu *et al.*, 2016). Interestingly, we note here that iPSC transfection appears to result in preferentially transfected cells near the outside of iPSC colonies, with a lack of transfection near the center of the colonies (Figure 4.8). This may be due to proliferation near the center of stem colonies resulting in newly forming untransfected cells in the center and older cells near the outside of colonies that have undergone transfection and are poised for differentiation. Additionally, HDR is typically reported as occurring with a frequency of less than 10% in the context of genome editing experiments (Anzalone *et al.*, 2019; Maruyama *et al.*, 2015; Mali *et al.*, 2013). In an effort to increase HDR, others have modified ssODN designs. Richardson *et al.* found that ssODNs that are oriented with 36 nucleotides PAM distal and 91 nucleotides PAM proximal and complementary to the non-target strand increased frequency of HDR in two transformed cell lines and at six different loci, resulting in frequencies of up to 57 ± 5% (2016). Additionally, a wide range of ssODN concentrations were tested. Despite adapting our ssODNs to adhere to these symmetry parameters and exploring alternate ssODN concentration, we still failed to detect HDR in patient-derived iPSCs (data not shown). CRISPR-Cas9 induced DSBs have also been reported to result in p53-mediated DNA damage response in human cells with an active p53

pathway, resulting in selective apoptosis (Haapaniemi *et al.*, 2018). Cells that remain in the G1 phase were also shown to have a decreased genome editing. These findings are a concern for CRISPR HDR, since they indicate that not only are healthy cells with a functioning p53 pathway less likely to be edited, but cells lacking functioning DNA repair may be preferentially selected for post-editing. Thus, this is a concern for propagating cells poised for oncogenesis.

In this chapter, two alternative genome editing strategies to target *GBA1* are utilized: base and prime editing. Base editors are an elegant combination of nCas9 linked to a deaminase which are capable of targeting a predictable five-nucleotide stretch of DNA to convert adenine to guanine or cytosine to thymine. Base editors are more efficient than CRISPR HDR but have more rigid design parameters. We find that the middle nucleotide within the five-nucleotide base editing window is more efficiently edited than other nucleotide positions, as evidenced by the majority of *HPRT1* editing (27 ± 4.3%) at the middle base. Only 3 ± 1.5% base editing was observed in position +5, and no editing in positions +1 or +2 across three trials. This may be a *HPRT1* specific finding; however, others have noted that base editing occurs at a greater frequency at the center of the base editing window, suggesting that centering the target base at position +3 will generate higher editing efficiencies across loci (Anzalone *et al.*, 2019). Further research into preferential base editor target sequence context has elucidated sequence-dependent activity. For example, editing efficiency has been found to be significantly lower when the 5' base of a target cytosine is a guanine (Rees and Liu, 2018; Komor *et al.*, 2016). In the case of base editing at *GBA1*, we find that across three cell types, editing of the c.1226 A>G mutation was obtained only in patient-derived FBs. In this genomic context, a bystander base was located in the center of the base editing window. Unsurprisingly, we find that this center base is edited in HEK293, FBs, and iPSCs with

an efficiency of 54%, 42%, and 4%, respectively. The bystander edit occurs within the 5' intron and may induce a splice site mutation within the intronic region. This putative splice site mutation is not reported in the literature and its impact on the resulting mRNA transcript should be clarified through cDNA analysis. This particular sequence context illuminates the aforementioned unfavourable guanine nucleotide 5' of the target cytosine. The target base, c.1226 G (base editing target is cytosine on the antisense strand), is located at the 5' end of the base editing window with a guanine immediately 5' of this site. Lastly, editing efficiencies across cell types are associated with transfection efficiencies, with the highest electroporation efficiency in HEK293 and the lowest electroporation efficiency in human iPSCs (Figure 4.11 –.13). Further optimization of electroporation in human iPSCs would be expected to increase the overall base editing efficiency.

Base editing strategies often lack design flexibility. gRNA testing and optimization is almost never an option given the narrow base editing window. To mitigate this issue, combined with the aforementioned inherent issues with the CRISPR HDR system, prime editing has been proposed as a more versatile alternative (Anzalone *et al.*, 2019). Prime editors combine the design flexibility associated with CRISPR HDR, but do not introduce DSBs. Here, we tested the prime editing system in HEK293, fibroblasts and iPSCs using a validated pegRNA that introduces a CTT insertion in the *HEK3* locus to optimize on-target efficiency (Anzalone *et al.*, 2019). We again find that installation efficiency is tied to electroporation efficiency. The highest installation efficiencies were seen in HEK293 cells (12.5 ± 0.5%), followed by iPSCs (5.1%), and fibroblasts (0%), with electroporation efficiencies of 82.3 ± 2.9%, 38.7 ± 6.3%, and 22.5 ± 2.4%, respectively. We find that culturing transfected cells for 96 hours post-transfection results in the highest editing efficiency (12.5 ± 0.5%) when compared to 72 hours (10.8 ± 0.2%) or 24 - 48 hours (no detectable

editing, data not shown). Anzalone *et al.* report ~70% CTT insertion at the *HEK3* locus; however, they use a more efficient version of prime editing (PE3) that requires a secondary gRNA directed to the non-edited strand, thus directing DNA repair systems to the non-edited strand, resulting in retention of the edit (2019). We opted not to use the PE3 system here because the addition of a second gRNA would increase the total required plasmid vector volume to exceed 10% of the total electroporation reaction volume. We have noted that increasing the total volume of delivered cargo beyond 10% drastically decreases electroporation efficiency and cell survival (unpublished finding).

In order to overcome the abovementioned issues with base editing at this genomic location, a pegRNA targeting the *GBA1* c.1226 A>G mutation was designed. This pegRNA incorporates a silent mutation in the PAM, thus reducing the ability of Cas9 retargeting and introducing an additional marker to determine successful genome editing at this target site. The 5' spacer region of this pegRNA does not overlap c.1226, but instead directs Cas9 cleavage seven nucleotides downstream from c.1226. This pegRNA can therefore be trialed in cells with the wildtype c.1226 A genotype, such as HEK293 cells. gDNA post-transfection can be screened for PAM site alterations as a preliminary assessment of on-targeting efficiency. Here, preliminary results indicate 1% prime editing of the PAM silent mutation in *GBA1* exon 9 in HEK293 cells, although going forward these findings must be further validated by performing several replicates and confirmed as positive editing events through additional *in silico* assays, such as TIDE analyses or NGS. Further validation of these results will be carried out in patient-derived FBs to determine if c.1226 A>G editing is occurring along with the silently mutated PAM. Electroporation efficiency will continue to be optimized by testing alternate electroporation parameters. Additionally, PE3b,

the PE3 system with a gRNA targeting the non-edited strand at the location where the edit has been introduced, will be designed and tested in an attempt to elicit editing at this site (Anzalone *et al.*, 2019). We are also exploring the benefits of G0/G1 cell cycle synchronization to achieve higher editing efficiencies (Chang *et al.*, 2019).

4.5. Conclusions

The base editing system is a suitable genome editing technique for editing genomic sites in HEK293 with high efficiency. Prime editing installation efficiencies at a control locus in HEK293 are similar to efficiencies expected (Anzalone *et al.*, 2019). For both the base editing and prime editing systems, we observed a very low editing or installation efficiency in human iPSCs, consistently <5%. It is not surprising that installation efficiencies tend to decrease with decreased transfection efficiency. An extended culture time beyond 72 hours post-transfection is crucial in obtaining measurable installation and editing efficiencies with both systems. The results of *HPRT1* and *GBA1* base editing experiments also suggest that the 3rd base of the base editing window is the most amenable to editing. This is evidenced by *GBA1* base editing in iPSCs occurring only at the 3rd position bystander base, and the greatest editing efficiencies in *HPRT1* being that of the target base in the 3rd position base editing window as well. This finding is supported by the literature (Anzalone *et al.*, 2019; Gaudelli *et al.*, 2017). CRISPR HDR, base editing and prime editing systems all contribute to the existing suite of genome editing strategies. The optimal system for a desired edit is determined by a number of sequence-context factors and the ultimate desired editing outcome.

4.6. Supplementary information

Supplementary Table 4.1. Forward and reverse primer sequences for PCR amplification of *NAGLU* exon 2, 5, and 6, *GBA1* exon 9, *HPRT1* and *HEK3*. All thermal cycling protocols followed the following structure: 94°C for 3 minutes initial denaturing, 35 or 37 cycles of 94°C for 30 seconds, annealing temperature for 30 seconds, and 72°C for 1 minute, followed by a final elongation at 72°C for 5 minutes.

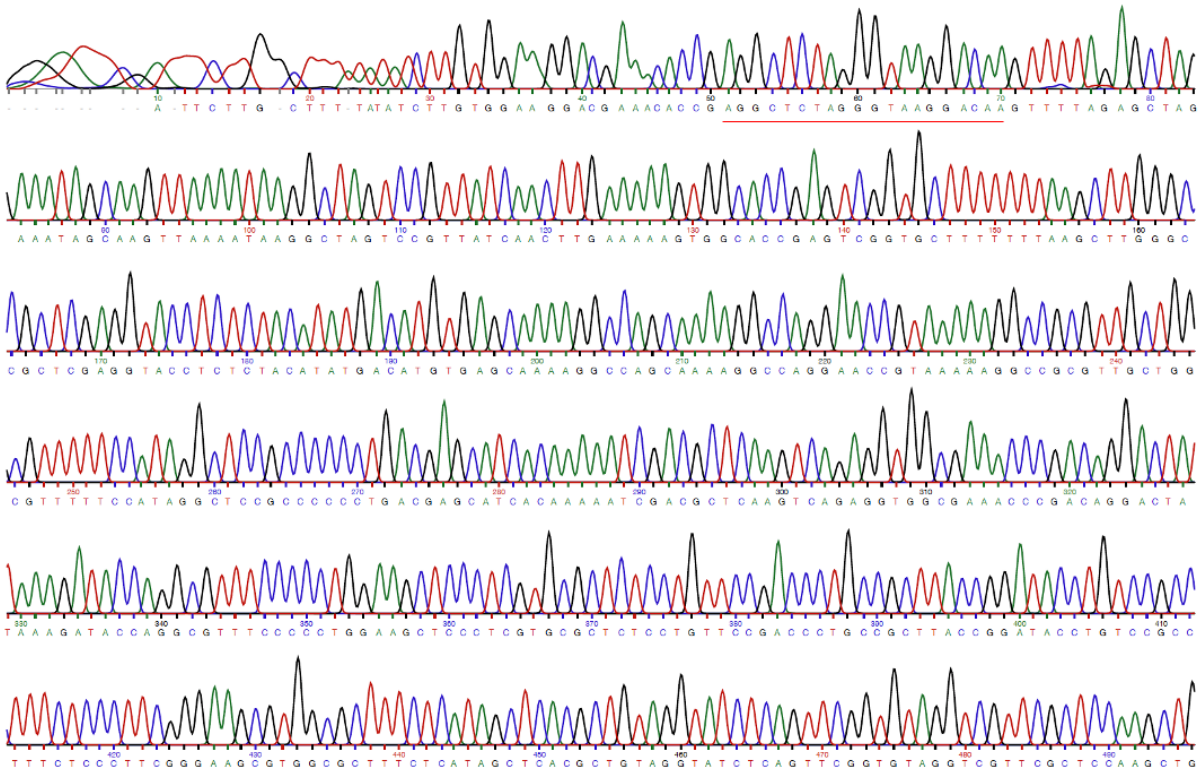
Application	Genomic Location	Amplicon (size bp)	Forward sequence (5' → 3')	Reverse sequence (5' →3')	Ta (°C)	Cycles
T7E1 gRNA1-3	*NAGLU* exon 5	6F/6R (434)	AAACCAGGAGCTGTAGAGAAGT	CTGCCTACCCCTACTGACATCT	55	37
T7E1 gRNAX,Y,C,U	*NAGLU* exon 2	3F/3R (595)	CCCTGCCCATCTGTTAGACT	GCACGTTGAAAGCACTTCTA	55	37
REGN-RFLP gRNAX, Y, C, U	*NAGLU* exon 2	2F/3R (379)	AGAAGGGCCGAGTTTGGAG	GCACGTTGAAAGCACTTCTA	51	37
RGEN gRNA+33-+70	*GBA1* exon 9	E8.5/S10 (1350)	TGTGCAAGGTCCAGGATCAG	TTCAGCCCACTTCCCAGACC	52	35
RGEN gRNAA, B, G	*NAGLU* exon 6	7F/7R (444)	GGCCCTCTGTTTCATCACTC	AAATCTGGCACTGGGTCCTT	55	37
Base editing	*GBA1* exon 9	E8.5/P2II (749)	TGTGCAAGGTCCAGGATCAG	GACAAAGTTACGCACCCAAT	51	37
Prime editing	*GBA1* exon 9	8.4/P2II (357)	CCATGATTCCCTATCTTCCC	GACAAAGTTACGCACCCAAT	56	35
Base editing	*HPRT1*	BEHPRTFWD/ BEHPRTRV (304)	TTTCTGTAGGACTGAACGTCTTGCTC	ATCTCACTGTAACCAAGTGAAATGAAAGC	57	35
Prime editing	HEK3	HEK3F/HEK3R (296)	ATGTGGGCTGCCTAGAAAGG	CCCAGCCAAACTTGTCAACC	52	35
*Hpy*CH4III digest	*NAGLU* exon 5	MAF/MAR (142)	ATCATCGGGAGCCTCTTCCT	CCTCATAGACGGCAGTGGT	51	37

Supplementary Table 4.2. gRNA sequences and associated PAM sequences for CRISPR-Cas9. The modified nucleotide to match the mutant base is underlined.

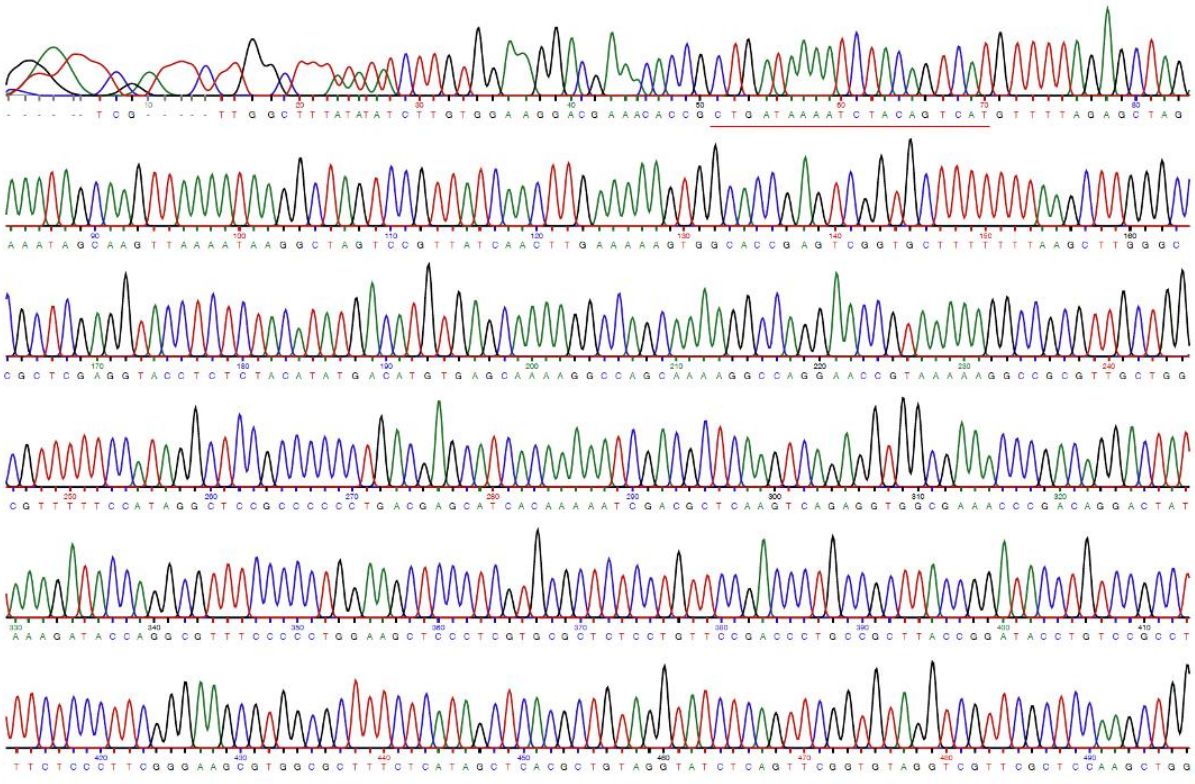
sgRNA	Sequence (5' → 3')	PAM
gRNA1	TGAGAGCTGATCAAAGAGTT	TGG
gRNA2	GGCACAGACCACATCTAT	TGG
gRNA3	ACTCTTTGATCAGCTCTCAC	AGG
gRNAA	AGCACCAGCTGCAGTTCTGG	GGG
gRNAB	CGGGCCCCCAGAACTGCAGC	TGG
gRNAG	CGCCCAGATCAGGGCTGTGC	TGG
gRNAY	CTGGGAGCGAAAGATAGACT	TGG
gRNAX	GGCCCAGTCCCACCACACGA	AGG
gRNAU	GGAGCGGCCAGGAGGCCATC	AGG
gRNAC	GGCGCTGAATGGCATCAACC	TGG
gRNA+33	GTTCAGGGCAAGGTTCCAGT	CGG
gRNA+49	TGGGTCCTCCTTCGGGGTTC	AGG
gRNA+58	CACCCAATTGGGTCCTCCTT	CGG
gRNA+70	GACAAAGTTACGCACCCAAT	TGG

Supplementary Table 4.3. gRNA sequences ligated into MLM3636 (Addgene #43860) and associated PAM sequences for base editing. Italicized bases indicate additional required guanine for gRNA expression; lowercase characters denote require overhangs for oligonucleotide annealing into MLM3636; uppercase characters denote gRNA sequence (Kwart *et al.*, 2017).

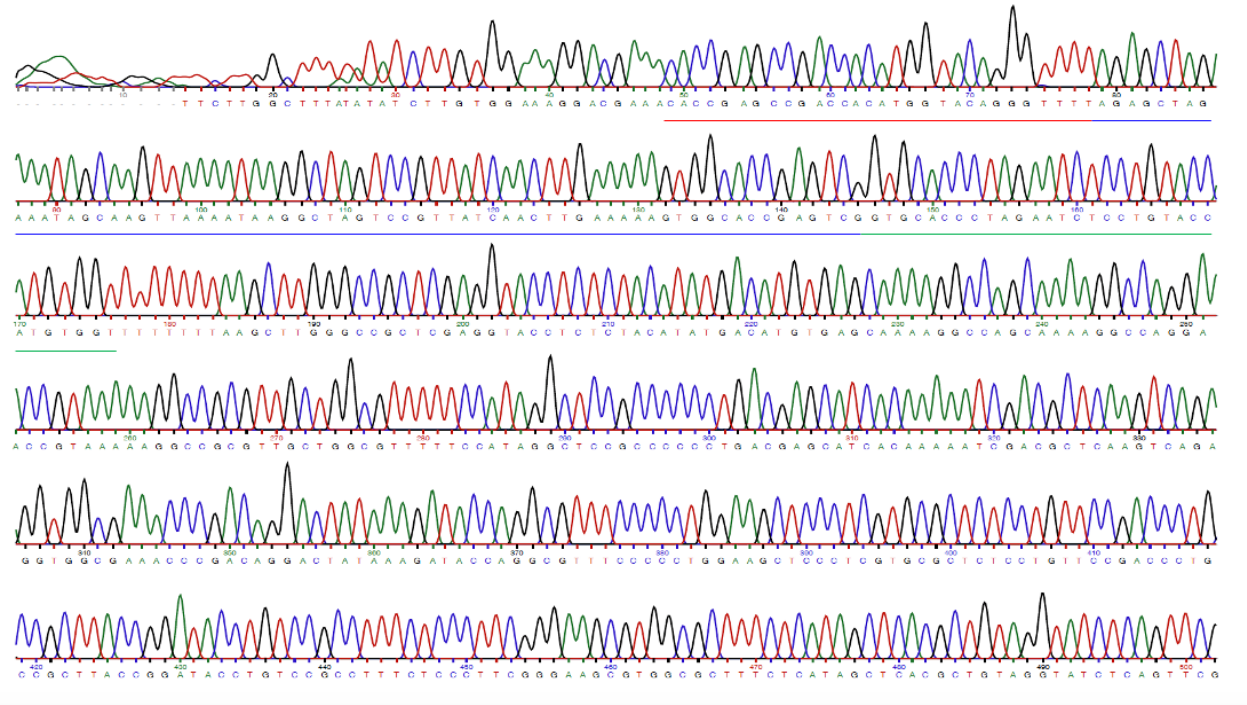
sgRNA	Forward Oligo (5' → 3')	Reverse Oligo (5' → 3')	PAM
BE_*HPRT1*	acacc*g*CTGATAAAATCTACAGTCATg	aaaacATGACTGTAGATTTTATCAG*c*g	AGG
BE_*GBA1*	acacc*g*AGGCTCTAGGGTAAGGACAA	aaaacTTGTCCTTACCCTAGAGCCT*c*g	AGG



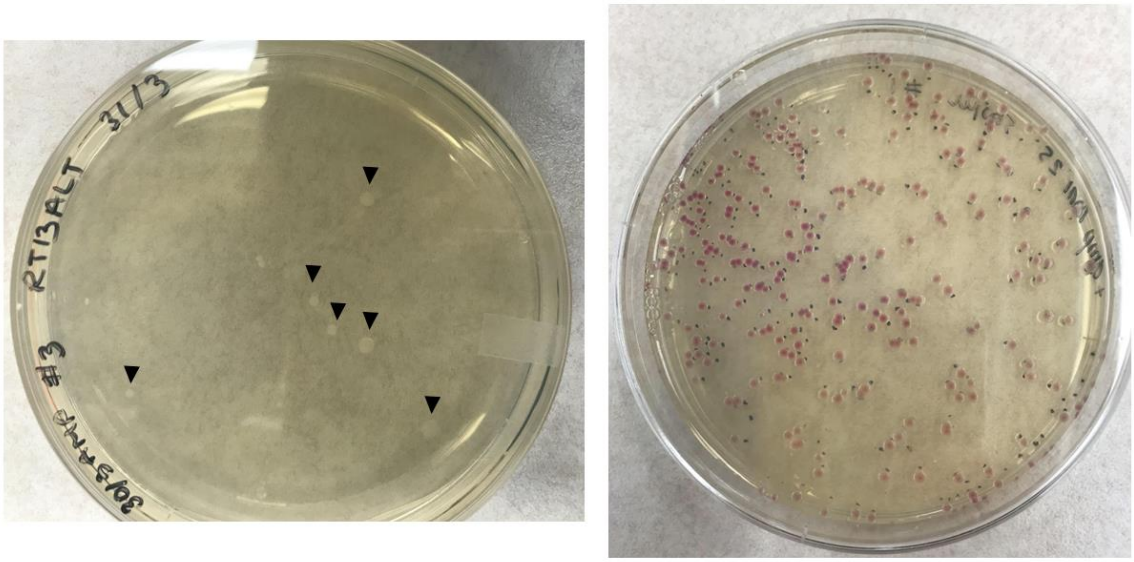
Supplementary Figure 4.1. Sanger sequencing DNA chromatogram showing insertion of *GBA1* c.1226 A>G targeting gRNA (red) into MLM3636 (Addgene #43860).



Supplementary Figure 4.2. Sanger sequencing DNA chromatogram showing insertion of *HPRT1* targeting gRNA (red) into MLM3636 (Addgene #43860).



Supplementary Figure 4.3. Sanger sequencing DNA chromatogram showing insertion of *GBA1* c.1226 A>G targeting pegRNA into pU6-pegRNA-GG-acceptor (Addgene #132777). Red is the spacer, blue is the scaffold, and green is the RT-primer binding site.



Supplementary Figure 4.4. DH5α Competent Cells transformed with pU6-pegRNA-GG-acceptor (Addgene #132777) (right) and pU6-pegRNA-GG-acceptor (Addgene #132777) ligated with c.1226 A>G targeting pegRNA (left).

5. Concluding Remarks

5.1. Purpose

The standard of care for GD and MPS IIIB patients with CNS involvement focuses on treatment of symptoms as they arise, followed by palliative care (Shapiro *et al.*, 2019). A number of contributing factors make these LDs attractive candidates for cell and gene therapies: (1) they currently lack suitable treatments to increase the overall standard of care, (2) small increases in enzyme levels may offer therapeutic benefit, and (3) functional lysosomal enzymes are trafficked to neighbouring cells resulting in cross-correction (Bunge *et al.*, 1998; Fratantoni *et al.*, 1968; Kaplan *et al.*, 1977; Sands and Davidson, 2006). Recent advances in genome editing technologies and stem cell biology have paved the way for developing cell regenerative therapies and gene therapies for the treatment of genetic diseases. Evidence indicates that early intervention is important in ensuring best possible outcomes for GD and MPS IIIB patients (Gilkes *et al.*, 2016; Lee *et al.*, 2019; Enderlin *et al.*, 2003). The rarity of LDs can often lead to a diagnostic odyssey, resulting in delays in interventions and poor outcome (Brady *et al.*, 2013). The purpose of this dissertation was to address a gap in suitable treatment options for patients with GD and MPS IIIB, by exploring genome editing techniques to target patient-specific mutations in *GBA1* and *NAGLU* and adapt existing genetic diagnostic techniques to identify mutations in these genes as well.

5.2. Identification of a novel compound heterozygous mutation in MPS IIIB and screening applications for LD mutations

More than 150 mutations have been identified in *NAGLU* to date, many of which are specific to single patients and their families (Khorrami *et al.*, 2019). Through PCR amplification

followed by Sanger sequencing, a novel compound heterozygous mutation was identified in *NAGLU*. Skin FBs were found to have 0.6% of normal NAGLU enzyme activity and ~50% NAGLU protein upon immunoblotting. The two identified mutations, p.Y140C and p.R297X, are located in exons 2 and 5, respectively, and are predicted to result in a non-functional NAGLU protein and either an unstable mRNA transcript or a truncated protein that is rapidly degraded. Additionally, RFLP-based methods for identifying p.R297X and p.E153K *NAGLU* mutations are demonstrated using mismatch PCR amplification followed by restriction digest. Lastly, a ddPCR was adapted to identify the common p.N370S mutation in *GBA1* using a functional gene-specific PCR method, followed by spiking p.N370S DNA in wildtype DNA. This ddPCR method shows specificity for p.N370S alleles but requires further optimization for accurate detection of this common mutation against wildtype alleles, such as in the setting of cffDNA obtained from maternal plasma.

5.3. Reprogramming MPS IIIB and GD patient fibroblasts to pluripotency

The initial discovery of a set of four transcription factors, c-Myc, Klf4, Oct3/4, and Sox2, that allowed for derivation of pluripotent stem cells from terminally differentiated cells has elevated the ability to model disease and has progressed regenerative therapies significantly (Takahashi and Yamanaka, 2006). iPSCs are especially important in modeling neurodegenerative disease, such as GD and MPS IIIB, due to their ability to be differentiated to NPCs (Kozhich *et al.*, 2013). A non-integrating SeV reprogramming method was used to obtain *GBA1*$^{N370S/N370S}$ and *NAGLU*$^{E153K/E153K}$ iPSCs with a 0.2% and 0.3% efficiency, respectively. These reprogramming efficiencies fall within the range reported in the literature (Schlaeger *et al.*, 2015; Huangfu *et al.*, 2008). Pluripotency was confirmed through APLS, flow cytometry for differentiation marker

SSEA-1 and pluripotency marker SSEA-4, and trilineage staining for endoderm marker GATA-4, mesoderm marker brachyury, and ectoderm markers Otx-2 and SOX1, and TaqMan™ hPSC Scorecard™ qRT-PCR assays. Sanger sequencing confirmed that the homozygous *GBA1* and *NAGLU* genotypes were maintained post-reprogramming. Increased cell death and spontaneous differentiation to EBs was observed in early passage MPS IIIB iPSCs. It was hypothesized that this was caused by partially degraded or improperly sulfated HSPGs either unable to interact with or sequestering FGF2, preventing interactions at FGF receptors and impairing cell proliferation. Supplementation of 100 ng/mL FGF2 in MPS IIIB iPSC culture media resulted in a significant increase in cell confluency after 48 hours and persisting through 96 hours ($n = 5$, $p \leq 0.05$, Kruskal-Wallis tests with post hoc Bonferroni procedure and simultaneous Mann-Whitney U paired tests), but not in GD or healthy human iPSCs. Future studies should be conducted to further elucidate strategies to overcome pluripotency issues caused by issues in HS catabolism. GD and MPS IIIB iPSCs can be used to model neurodegeneration by differentiating these cells to neural precursors and neurons, and isogenic control cell lines should be created using genome editing methodologies.

5.4. Genome editing outcomes in *GBA1* and *NAGLU*

Genome editing strategies are of particular interest for developing models and therapeutic strategies for LDs. Evidence for secondary GoF roles of not only mutant *GBA1*, but also *NAGLU*, underscore the importance of targeting the existing gene for gene correction over safe harbour knock-in strategies (Sidransky *et al.*, 2009; Chinen *et al.*, 2005). In this chapter, gRNAs were designed to target three different pathogenic mutations in *NAGLU* (p.E153K, p.R297X, p.Y140C) and one in *GBA1* (p.N370S). Of the gRNAs tested, ten of 11 show more than 70% on-target Cas9

cleavage via RGEN-RFLP assay. One of these gRNAs yielded >20% indels upon T7E1 analysis. Despite the aforementioned gRNA characterization and optimization of ssODN concentration and homology arm size and symmetry about the Cas9 cut site, CRISPR HDR was not detectable. Instead, a base editing strategy to target patient-specific mutations was used. This base editing system was tested by targeting the *HPRT1* gene in HEK293 cells. A transfection efficiency of 51.7 ± 15.5% and an editing efficiency of 6 ± 0.5% were obtained. Editing efficiency is the highest for the middle nucleotide position (27 ± 4.3%) of the base editing window, which fits with recent literature findings that show higher editing efficiencies at this middle nucleotide (Anzalone *et al.*, 2020). A base editing strategy targeting the *GBA1* p.N370S mutation was also explored in patient-derived iPSCs, skin FBs, and HEK293 cells. This p.N370S-targeting base editing strategy shows an editing efficiency of 4%, 42%, and 54%, respectively, in these cell types. Base editing in HEK293 cells showed up to a 95% editing efficiency when cultured for 120 hours prior to genomic DNA extraction. Preliminary prime editing results are also reported, demonstrating an installation efficiency of a three bp (CTT) insertion in the *HEK3* locus of 11.8% and 5.1% in HEK293 cells and iPSCs. No detectable CTT installation in *HEK3* was measured in skin FBs. Installation efficiencies increased with increasing transfection efficiencies, and thus further optimization of transfection should be explored.

5.5. Future directions

Although the *NAGLU* p.R297X mutation has been previously characterized for its impact on enzyme activity and residual protein levels, p.E153K and p.Y140C have yet to be characterized (Meijer *et al.*, 2017). Going forward, it will be important to determine the impact of these mutations on residual protein levels as this may influence the direction of suitable therapies, such as PCT

which requires residual enzyme to chaperone (Parenti *et al.*, 2015). Further elucidation of the impact of these mutations on overall enzyme function can help predict the anticipated disease severity resulting from these mutations either in homozygosity or heterozygosity with another mutation. Although inherent restriction sites which incorporate the p.Y140C mutation do not exist, site-directed mutagenesis as was demonstrated here to create restriction sites incorporating the p.E153K and p.R297X mutations could be used to also identify p.Y140C.

ddPCR is a promising technique for NIPD of monogenic diseases. The p.N370S and reference probes designed and tested here, plus the nested PCR and massive dilution strategy should be tested on clinical plasma samples containing cffDNA from pregnancies at risk for GD and against those known to be unaffected to determine the sensitivity of this technique. A comprehensive TaqMan™ panel assay encompassing a set of SNPs with high minor allele frequencies should be utilized to quantify the amount of cffDNA present in the maternal plasma sample, such as is described in Camunas-Soler *et al.* (2018). cffDNA fragments tend to be shorter in length than maternal circulating cell-free DNA, thus methodologies of cffDNA enrichment via size selection may increase the total proportion of cffDNA for ddPCR analysis and negate the current NIPT threshold of ~4% total fetal fraction (Hu *et al.*, 2019). Additionally, ddPCR probes have been designed to target the p.L444P mutation in *GBA1* to be tested using this aforementioned ddPCR technique. ddPCR probes that target other common *GBA1* mutations resulting in severe phenotypes, such as the complex allele Rec*Nci*I, and splicing variant IVS2 +1, should also be validated. Furthermore, the suitability of a ddPCR-based NIPT panel to test for GD in at-risk pregnancies should be explored that includes ddPCR assays for all of these aforementioned common GD mutations.

Further investigations into the impact of partially degraded and undegraded HS in not only MPS IIIB iPSC proliferation and pluripotency, but also other disorders of HS catabolism, should focus on determining if (1) HS accumulation results in the sequestration of FGF2 at the cell surface, or if (2) partially degraded HS fails to bind FGF2, thus reducing FGF2 half-life. Building on existing studies that have shown supplemental HS to increase the half-life of FGF2 in healthy iPS cells, supplemental HS should be added to MPS IIIB iPS cultures at a starting concentration of 5 μg/mL (Horiguchi *et al.*, 2018). Cultures should be analysed for FGF2 concentration, cell pluripotency and differentiation markers, and growth rate. Additionally, increased NAGLU enzyme activity and total enzyme levels were recently reported in MPS IIIB skin FBs cultured under hypothermic conditions (Meijer *et al.*, 2018). Future work should focus on assessing the impact of culturing MPS IIIB skin FBs throughout the reprogramming process under hypothermic conditions to determine if increased NAGLU enzyme activity persists throughout reprogramming and influences overall reprogramming efficiency and maintenance of pluripotency during early iPSC emergence. Since culturing MPS IIIB cells under hypothermic conditions is predicted to improve NAGLU folding, MPS IIIB genotypes that result in a non-functional protein product are potential candidates for this method (Meijer *et al.*, 2018).

Genome editing strategies have progressed significantly since the beginning of this doctoral dissertation work. The development of base and prime editing strategies has greatly increased the scope of applicable target mutations for editing and elevated editing efficiencies in disease-relevant cell types. Through the base editing of *HPRT1* and *GBA1* demonstrated here, we show evidence of increased editing efficiencies at the 3rd nucleotide within the base editing

window, which supports recent findings (Anzalone *et al.*, 2020). Since genome editing efficiencies greatly differ depending on the surrounding nucleic acid sequence to the target site, GC content, epigenetic factors, and nearby alternative PAM sequences, precise genomic nuances that impact on-target efficiency will need to be addressed (Daer *et al.*, 2017; Isaac *et al.*, 2016; Cong *et al.*, 2013). Specific to this work, base and prime editing strategies targeting the previously discussed p.R297X, p.E153K, and p.Y140C mutations in *NAGLU*, and p.L444P, Rec*Nci*I and IVS2 +1 mutations in *GBA1* should be designed and validated. The most suitable genome editing strategy should be determined depending on genomic location and sequence characteristics proximal to these mutations of interest. Additionally, given the presence of a bystander base within in the base editing window of our described *GBA1* base editing strategy at the favourable third nucleotide position which is predicted to create a splice site alteration, next steps should focus on (1) establishing a prime editing method to circumvent this bystander edit, or (2) establishing a screening strategy to identify cells that have undergone only p.N370S mutation editing. Furthermore, edited patient-derived cells should be characterized for off-target effects, GCase activity and glucosylceramide levels, in addition to α-synuclein levels in comparison to isogenic GD cells. Going forward, validated base and prime editing strategies *in vitro* will be further assessed in murine models of GD and MPS IIIB, where optimal *in vivo* genome editing construct delivery strategies and appropriate dosing studies should be conducted, followed by assessment of therapeutic outcome in these animals.

5.6. Significance

Three overarching themes are presented in this dissertation: (1) RFLP and ddPCR strategies can be leveraged for identification of MPS IIIB and GD mutations from patients' cells; (2) inherent disease pathology is an imperative consideration when utilizing iPSCs for disease modelling and the development the therapeutic strategies, and (3) alternative genome editing strategies such as base and prime editing offer more efficient solutions to correcting patient-specific mutations *in vitro* than CRISPR HDR. To our knowledge, this is the first investigation into genome editing strategies for pathogenic mutation correction in *NAGLU* and the first report of base editing in *GBA1*. Going forward, multiple genome editing strategies should be compared when pursuing genome editing for correction of patient-specific mutations. Furthermore, this work shows that base editing is a highly efficient editing strategy when the target nucleotide is centered in the base editing window and alternative bystander bases are not present, and is a recommended genome editing strategy for experiments that satisfy these parameters. The findings presented here will translate to the future development of cell and gene therapies in regenerative medicine.

6. Bibliography

Aflaki, E., Westbroek, W. & Sidransky, E. The Complicated Relationship between Gaucher Disease and Parkinsonism: Insights from a Rare Disease. *Neuron* **93**, 737-746 (2017).

Al-Jasmi, F.A. *et al.* Prevalence and Novel Mutations of Lysosomal Storage Disorders in United Arab Emirates : LSD in UAE. *JIMD Rep* **10**, 1-9 (2013).

Aldenhoven, M. *et al.* Quality of life of Hurler syndrome patients after successful hematopoietic stem cell transplantation. *Blood Adv* **1**, 2236-2242 (2017).

Allende, M.L. *et al.* Cerebral organoids derived from Sandhoff disease-induced pluripotent stem cells exhibit impaired neurodifferentiation. *J Lipid Res* **59**, 550-563 (2018).

Altarescu, G. *et al.* PGD on a recombinant allele: crossover between the TSC2 gene and 'linked' markers impairs accurate diagnosis. *Prenat Diagn* **28**, 929-933 (2008).

Andrade, F., Aldámiz-Echevarría, L., Llarena, M. & Couce, M.L. Sanfilippo syndrome: Overall review. *Pediatr Int* **57**, 331-338 (2015).

Andria, G., Di Natale, P., Del Giudice, E., Strisciuglio, P. & Murino, P. Sanfilippo B syndrome (MPS III B): mild and severe forms within the same sibship. *Clin Genet* **15**, 500-504 (1979).

Antonarakis, S.E., Krawczak, M. & Cooper, D.N. Disease-causing mutations in the human genome. *Eur J Pediatr* **159 Suppl 3**, S173-178 (2000).

Anzalone, A.V., Koblan, L.W. & Liu, D.R. Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. *Nat Biotechnol* **38**, 824-844 (2020).

Anzalone, A.V. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149-157 (2019).

Armand, P. *et al.* Validation and refinement of the Disease Risk Index for allogeneic stem cell transplantation. *Blood* **123**, 3664-3671 (2014).

Ascending Dose Study of Genome Editing by the Zinc Finger Nuclease (ZFN) Therapeutic SB-318 in Subjects With MPS I. Available online: https://clinicaltrials.gov/ct2/show/NCT02702115. Accessed: 10 June 2019.

Ascending Dose Study of Genome Editing by the Zinc Finger Nuclease (ZFN) Therapeutic SB-913 in Subjects With MPS II. Available online: https://clinicaltrials.gov/ct2/show/NCT03041324. Accessed: 10 June 2019.

Ashoor, G., Syngelaki, A., Poon, L.C., Rezende, J.C. & Nicolaides, K.H. Fetal fraction in maternal plasma cell-free DNA at 11-13 weeks' gestation: relation to maternal and fetal characteristics. *Ultrasound Obstet Gynecol* **41**, 26-32 (2013).

Babon, J.J., McKenzie, M. & Cotton, R.G. The use of resolvases T4 endonuclease VII and T7 endonuclease I in mutation detection. *Mol Biotechnol* **23**, 73-81 (2003).

Bach, F.H., Albertini, R.J., Joo, P., Anderson, J.L. & Bortin, M.M. Bone-marrow transplantation in a patient with the Wiskott-Aldrich syndrome. *Lancet* **2**, 1364-1366 (1968).

Baehner, F. *et al.* Cumulative incidence rates of the mucopolysaccharidoses in Germany. *J Inherit Metab Dis* **28**, 1011-1017 (2005).

Balasubramanian, V. & Ramanathan, M. Glycosaminoglycans alter the conformation of interferon-gamma. *Cytokine* **12**, 466-471 (2000).

Balwani, M., Grace, M.E. & Desnick, R.J. Gaucher disease: when molecular testing and clinical presentation disagree -the novel c.1226A>G(p.N370S)--RecNcil allele. *J Inherit Metab Dis* **34**, 789-793 (2011).

Ban, H. *et al.* Efficient generation of transgene-free human induced pluripotent stem cells (iPSCs) by temperature-sensitive Sendai virus vectors. *Proc Natl Acad Sci U S A* **108**, 14234-14239 (2011).

Barnes, J. *et al.* Modeling the neuropsychiatric manifestations of Lowe syndrome using induced pluripotent stem cells: defective F-actin polymerization and WAVE-1 expression in neuronal cells. *Mol Autism* **9**, 44 (2018).

Barrett, A.N., McDonnell, T.C., Chan, K.C. & Chitty, L.S. Digital PCR analysis of maternal plasma for noninvasive detection of sickle cell anemia. *Clin Chem* **58**, 1026-1032 (2012).

Barton, N.W., Furbish, F.S., Murray, G.J., Garfield, M. & Brady, R.O. Therapeutic response to intravenous infusions of glucocerebrosidase in a patient with Gaucher disease. *Proc Natl Acad Sci U S A* **87**, 1913-1916 (1990).

Beesley, C.E., Young, E.P., Vellodi, A. & Winchester, B.G. Identification of 12 novel mutations in the alpha-N-acetylglucosaminidase gene in 14 patients with Sanfilippo syndrome type B (mucopolysaccharidosis type IIIB). *J Med Genet* **35**, 910-914 (1998).

Bekris, L.M., Mata, I.F. & Zabetian, C.P. The genetics of Parkinson disease. *J Geriatr Psychiatry Neurol* **23**, 228-242 (2010).

Belfort, M. & Bonocora, R.P. Homing endonucleases: from genetic anomalies to programmable genomic clippers. *Methods Mol Biol* **1123**, 1-26 (2014).

Bembi, B. *et al.* Gaucher's disease with Parkinson's disease: clinical and pathological aspects. *Neurology* **61**, 99-101 (2003).

Benchoua, A. & Onteniente, B. Intracerebral transplantation for neurological disorders. Lessons from developmental, experimental, and clinical studies. *Front Cell Neurosci* **6**, 2 (2011).

Bennett, L.L. & Mohan, D. Gaucher disease and its treatment options. *Ann Pharmacother* **47**, 1182-1193 (2013).

Benraiss, A. *et al.* Human glia can both induce and rescue aspects of disease phenotype in Huntington disease. *Nat Commun* **7**, 11758 (2016).

Beutler, E., Gelbart, T. & West, C. The facile detection of the nt 1226 mutation of glucocerebrosidase by 'mismatched' PCR. *Clin Chim Acta* **194**, 161-166 (1990).

Bhutada, E., Pyragius, T., Petersen, S.G., Niemann, F. & Matsika, A. Perinatal Lethal Gaucher Disease due to RecNcil Recombinant Mutation in the. *Case Rep Pathol* **2018**, 2549451 (2018).

Biffi, A. Gene therapy for lysosomal storage disorders: a good start. *Hum Mol Genet* **25**, R65-75 (2016).

Biffi, A. Hematopoietic Stem Cell Gene Therapy for Storage Disease: Current and New Indications. *Mol Ther* **25**, 1155-1162 (2017).

Birrane, G. *et al.* Structural characterization of the α-N-acetylglucosaminidase, a key enzyme in the pathogenesis of Sanfilippo syndrome B. *J Struct Biol* **205**, 65-71 (2019).

Bishop, A.J. & Schiestl, R.H. Homologous recombination as a mechanism for genome rearrangements: environmental and genetic effects. *Hum Mol Genet* **9**, 2427-2334 (2000).

Bitton, A., Etzell, J., Grenert, J.P. & Wang, E. Erythrophagocytosis in Gaucher cells. *Arch Pathol Lab Med* **128**, 1191-1192 (2004).

Bjørlykke, Y. *et al.* Reprogrammed Cells Display Distinct Proteomic Signatures Associated with Colony Morphology Variability. *Stem Cells Int* **2019**, 8036035 (2019).

Blanc, V. & Davidson, N.O. APOBEC-1-mediated RNA editing. *Wiley Interdiscip Rev Syst Biol Med* **2**, 594-602 (2010).

Boulting, G.L. *et al.* A functionally characterized test set of human induced pluripotent stem cells. *Nat Biotechnol* **29**, 279-286 (2011).

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

Brady, J., Trehan, A., Landis, D. & Toro, C. Mucopolysaccharidosis type IIIB (MPS IIIB) masquerading as a behavioural disorder. *BMJ Case Rep* **2013** (2013).

Braunlin, E. *et al.* Enzyme replacement therapy for mucopolysaccharidosis VI: long-term cardiac effects of galsulfase (Naglazyme®) therapy. *J Inherit Metab Dis* **36**, 385-394 (2013).

Bruegmann, T., Deecke, K. & Fladung, M. Evaluating the Efficiency of gRNAs in CRISPR/Cas9 Mediated Genome Editing in Poplars. *Int J Mol Sci* **20** (2019).

Bultron, G. *et al.* The risk of Parkinson's disease in type 1 Gaucher disease. *J Inherit Metab Dis* **33**, 167-173 (2010).

Bunge, S. *et al.* Genotype-phenotype correlations in mucopolysaccharidosis type I using enzyme kinetics, immunoquantification and in vitro turnover studies. *Biochim Biophys Acta* **1407**, 249-256 (1998).

Burrow, T.A. *et al.* Prevalence and management of Gaucher disease, Pediatr Health Med Ther **2** (2011).

Buzzard, J.J., Gough, N.M., Crook, J.M. & Colman, A. Karyotype of human ES cells during extended culture. *Nat Biotechnol* **22**, 381-382; author reply 382 (2004).

Campbell, T.N. & Choy, F.Y. Gaucher disease and the synucleinopathies: refining the relationship. *Orphanet J Rare Dis* **7**, 12 (2012).

Camunas-Soler, J. *et al.* Noninvasive Prenatal Diagnosis of Single-Gene Disorders by Use of Droplet Digital PCR. *Clin Chem* **64**, 336-345 (2018).

Carstea, E.D. *et al.* Niemann-Pick C1 disease gene: homology to mediators of cholesterol homeostasis. *Science* **277**, 228-231 (1997).

Cavé, M.C. *et al.* Glycosaminoglycans bind human IL-27 and regulate its activity. *Eur J Immunol* (2020).

Chandrasegaran, S. & Carroll, D. Origins of Programmable Nucleases for Genome Engineering. *J Mol Biol* **428**, 963-989 (2016).

Chang, H.H.Y., Pannunzio, N.R., Adachi, N. & Lieber, M.R. Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat Rev Mol Cell Biol* **18**, 495-506 (2017).

Chang, M.Y. *et al.* One-step noninvasive prenatal testing (NIPT) for autosomal recessive homozygous point mutations using digital PCR. *Sci Rep* **8**, 2877 (2018).

Chang, M.Y. *et al.* Development of novel noninvasive prenatal testing protocol for whole autosomal recessive disease using picodroplet digital PCR. *Sci Rep* **6**, 37153 (2016).

Chang, Y.J. *et al.* CRISPR Base Editing in Induced Pluripotent Stem Cells. *Methods Mol Biol* **2045**, 337-346 (2019).

Chatterjee, P., Cheung, Y. & Liew, C. Transfecting and nucleofecting human induced pluripotent stem cells. *J Vis Exp* (2011).

Chavez, A. *et al.* Comparison of Cas9 activators in multiple species. *Nat Methods* **13**, 563-567 (2016).

Chen, B., Zou, W., Xu, H., Liang, Y. & Huang, B. Efficient labeling and imaging of protein-coding genes in living cells using CRISPR-Tag. *Nat Commun* **9**, 5065 (2018).

Chen, G. *et al.*, Thermal stability of fibroblast growth factor protein is a determinant factor in regulating self-renewal, differentiation, and reprogramming in human pluripotent stem cells. *Stem Cells* **30**, 623-630 (2013).

Chinen, Y., Tohma, T., Izumikawa, Y., Uehara, H. & Ohta, T. Sanfilippo type B syndrome: five patients with an R565P homozygous mutation in the alpha-N-acetylglucosaminidase gene from the Okinawa islands in Japan. *J Hum Genet* **50**, 357-359 (2005).

Chiti, F. & Dobson, C.M. Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. *Annu Rev Biochem* **86**, 27-68 (2017).

Cho, S.W. *et al.* Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res* **24**, 132-141 (2014).

Chow, R. D., Chen, J. S., Shen, J., & Chen, S. pegFinder: A pegRNA designer for CRISPR prime editing. bioRxiv 2020.05.06.081612; doi: https://doi.org/10.1101/2020.05.06.081612

Christensen, C.L., Ashmead, R.E. & Choy, F.Y.M. Identification and characterization of NAGLU mutations in a Sanfilippo B syndrome patient with a novel genotype (p.Y140C/p.R297X). *Archives of Clin and Biomed Res* **1**, *4* (2017).

Christensen, C.L., Ashmead, R.E. & Choy, F.Y.M. Cell and Gene Therapies for Mucopolysaccharidoses: Base Editing and Therapeutic Delivery to the CNS. *Diseases* **7** (2019).

Christensen, C.L. & Choy, F.Y.M. A Prospective Treatment Option for Lysosomal Storage Diseases: CRISPR/Cas9 Gene Editing Technology for Mutation Correction in Induced Pluripotent Stem Cells. *Diseases* **5** (2017).

Clark, W.T. *et al.* Assessment of predicted enzymatic activity of α-N-acetylglucosaminidase variants of unknown significance for CAGI 2016. *Hum Mutat* **40**, 1519-1529 (2019).

Cleary, M.A. & Wraith, J.E. Management of mucopolysaccharidosis type III. *Arch Dis Child* **69**, 403-406 (1993).

Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823 (2013).

Cosma, M.P. *et al.* The multiple sulfatase deficiency gene encodes an essential and limiting factor for the activity of sulfatases. *Cell* **113**, 445-456 (2003).

D'Avanzo, F., Rigon, L., Zanetti, A. & Tomanin, R. Mucopolysaccharidosis Type II: One Hundred Years of Research, Diagnosis, and Treatment. *Int J Mol Sci* **21** (2020).

Daer, R.M., Cutts, J.P., Brafman, D.A. & Haynes, K.A. The Impact of Chromatin Dynamics on Cas9-Mediated Genome Editing in Human Cells. *ACS Synth Biol* **6**, 428-438 (2017).

de Araujo, M.E.G., Liebscher, G., Hess, M.W. & Huber, L.A. Lysosomal size matters. *Traffic* **21**, 60-75 (2020).

de Duve, C., Pressman, B.C., Gianetto, R., Wattiux, R. & Appelmans, F. Tissue fractionation studies. Intracellular distribution patterns of enzymes in rat-liver tissue. *Biochem J* **60**, 604-617 (1955).

De Pasquale, V. & Pavone, L.M. Heparan sulfate proteoglycans: The sweet side of development turns sour in mucopolysaccharidoses. *Biochim Biophys Acta Mol Basis Dis* **1865**, 165539 (2019).

de Ru, M.H. *et al.* Enzyme replacement therapy and/or hematopoietic stem cell transplantation at diagnosis in patients with mucopolysaccharidosis type I: results of a European consensus procedure. *Orphanet J Rare Dis* **6**, 55 (2011).

de Ruijter, J. *et al.* Genistein in Sanfilippo disease: a randomized controlled crossover trial. *Ann Neurol* **71**, 110-120 (2012).

De Zio, D., Cianfanelli, V. & Cecconi, F. New insights into the link between DNA damage and apoptosis. *Antioxid Redox Signal* **19**, 559-571 (2013).

Debrand, E., Lykoudi, A., Bradshaw, E. & Allen, S.K. A Non-Invasive Droplet Digital PCR (ddPCR) Assay to Detect Paternal CFTR Mutations in the Cell-Free Fetal DNA (cffDNA) of Three Pregnancies at Risk of Cystic Fibrosis via Compound Heterozygosity. *PLoS One* **10**, e0142729 (2015).

Delgadillo, V., O'Callaghan, M.e.M., Gort, L., Coll, M.J. & Pineda, M. Natural history of Sanfilippo syndrome in Spain. *Orphanet J Rare Dis* **8**, 189 (2013).

Dersh, D., Iwamoto, Y. & Argon, Y. Tay-Sachs disease mutations in HEXA target the α chain of hexosaminidase A to endoplasmic reticulum-associated degradation. *Mol Biol Cell* **27**, 3813-3827 (2016).

Deuse, T. *et al.* Hepatocyte growth factor or vascular endothelial growth factor gene transfer maximizes mesenchymal stem cell-based myocardial salvage after acute myocardial infarction. *Circulation* **120**, S247-254 (2009).

Devkota, S. The road less traveled: strategies to enhance the frequency of homology-directed repair (HDR) for increased efficiency of CRISPR/Cas-mediated transgenesis. *BMB Rep* **51**, 437-443 (2018).

Dimitriou, E. *et al.* Gaucher disease: Biochemical and molecular findings in 141 patients diagnosed in Greece. *Mol Genet Metab Rep* **24**, 100614 (2020).

Doench, J.G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* **34**, 184-191 (2016).

Donati, M.A., Pasquini, E., Spada, M., Polo, G. & Burlina, A. Newborn screening in mucopolysaccharidoses. *Ital J Pediatr* **44**, 126 (2018).

Dorfman, A. & Matalon, R. The mucopolysaccharidoses (a review). *Proc Natl Acad Sci U S A* **73**, 630-637 (1976).

Dreyfuss, J.L. *et al.* Heparan sulfate proteoglycans: structure, protein interactions and cell signaling. *An Acad Bras Cienc* **81**, 409-429 (2009).

Duan, J. *et al.* Live imaging and tracking of genome regions in CRISPR/dCas9 knock-in mice. *Genome Biol* **19**, 192 (2018).

Eaton, S.L. *et al.* CRISPR/Cas9 mediated generation of an ovine model for infantile neuronal ceroid lipofuscinosis (CLN1 disease). *Sci Rep* **9**, 9891 (2019).

Emre, S. *et al.* Sanfilippo syndrome in Turkey: Identification of novel mutations in subtypes A and B. *Hum Mutat* **19**, 184-185 (2002).

Enderlin, C., Vogel, R. & Conaway, P. Gaucher disease. *Am J Nurs* **103**, 50-60; quiz 61 (2003).

Eng, C.M. *et al.* A phase 1/2 clinical trial of enzyme replacement in fabry disease: pharmacokinetic, substrate clearance, and safety studies. *Am J Hum Genet* **68**, 711-722 (2001).

Erwood, S. *et al.* Modeling Niemann-Pick disease type C in a human haploid cell line allows for patient variant characterization and clinical interpretation. *Genome Res* **29**, 2010-2019 (2019).

Fecarotta, S., Gasperini, S. & Parenti, G. New treatments for the mucopolysaccharidoses: from pathophysiology to therapy. *Ital J Pediatr* **44**, 124 (2018).

Fernandes, H.J. *et al.* ER Stress and Autophagic Perturbations Lead to Elevated Extracellular α-Synuclein in GBA-N370S Parkinson's iPSC-Derived Dopamine Neurons. *Stem Cell Reports* **6**, 342-356 (2016).

Fernandez-Botran, R., Yan, J. & Justus, D.E. Binding of interferon gamma by glycosaminoglycans: a strategy for localization and/or inhibition of its activity. *Cytokine* **11**, 313-325 (1999).

Fratantoni, J.C., Hall, C.W. & Neufeld, E.F. Hurler and Hunter syndromes: mutual correction of the defect in cultured fibroblasts. *Science* **162**, 570-572 (1968).

Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* **31**, 822-826 (2013).

Fukuhara, Y. *et al.* Histopathological and behavioral improvement of murine mucopolysaccharidosis type VII by intracerebral transplantation of neural stem cells. *Mol Ther* **13**, 548-555 (2006).

Fusaki, N., Ban, H., Nishiyama, A., Saeki, K. & Hasegawa, M. Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. *Proc Jpn Acad Ser B Phys Biol Sci* **85**, 348-362 (2009).

Gadola, S.D. *et al.* Impaired selection of invariant natural killer T cells in diverse mouse models of glycosphingolipid lysosomal storage diseases. *J Exp Med* **203**, 2293-2303 (2006).

Gaffke, L., Pierzynowska, K., Piotrowska, E. & Węgrzyn, G. How close are we to therapies for Sanfilippo disease? *Metab Brain Dis* **33**, 1-10 (2018).

Gaj, T., Gersbach, C.A. & Barbas, C.F. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol* **31**, 397-405 (2013).

Galvan, C., Camoletto, P.G., Cristofani, F., Van Veldhoven, P.P. & Ledesma, M.D. Anomalous surface distribution of glycosyl phosphatidyl inositol-anchored proteins in neurons lacking acid sphingomyelinase. *Mol Biol Cell* **19**, 509-522 (2008).

Garbade, S.F. *et al.* FDA orphan drug designations for lysosomal storage disorders - a cross-sectional analysis. *PLoS One* **15**, e0230898 (2020).

Gaudelli, N.M. *et al.* Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**, 464-471 (2017).

Germain, D.P. *et al.* Ten-year outcome of enzyme replacement therapy with agalsidase beta in patients with Fabry disease. *J Med Genet* **52**, 353-358 (2015).

Gilkes, J.A., Bloom, M.D. & Heldermon, C.D. Mucopolysaccharidosis IIIB confers enhanced neonatal intracranial transduction by AAV8 but not by 5, 9 or rh10. *Gene Ther* **23**, 263-271 (2016).

Ginns, E.I. *et al.* Gene mapping and leader polypeptide sequence of human glucocerebrosidase: implications for Gaucher disease. *Proc Natl Acad Sci U S A* **82**, 7101-7105 (1985).

Giugliani, R. *et al.* A dose-optimization trial of laronidase (Aldurazyme) in patients with mucopolysaccharidosis I. *Mol Genet Metab* **96**, 13-19 (2009).

Glick, D., Barth, S. & Macleod, K.F. Autophagy: cellular and molecular mechanisms. *J Pathol* **221**, 3-12 (2010).

Gratwohl, A. *et al.* Hematopoetic stem cell transplantation for solid tumors in Europe. *Ann Oncol* **15**, 653-660 (2004).

Guce, A.I. *et al.* Catalytic mechanism of human alpha-galactosidase. *J Biol Chem* **285**, 3625-3632 (2010).

Guggenbuhl, P., Grosbois, B. & Chalès, G. Gaucher disease. *Joint Bone Spine* **75**, 116-124 (2008).

Gupta, R.M. & Musunuru, K. Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9. *J Clin Invest* **124**, 4154-4161 (2014).

Guschin, D.Y. *et al.* A rapid and general assay for monitoring endogenous gene modification. *Methods Mol Biol* **649**, 247-256 (2010).

Gutierrez-Aranda, I. *et al.* Human induced pluripotent stem cells develop teratoma more efficiently and faster than human embryonic stem cells regardless the site of injection. *Stem Cells* **28**, 1568-1570 (2010).

Haapaniemi, E., Botla, S., Persson, J., Schmierer, B. & Taipale, J. CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. *Nat Med* **24**, 927-930 (2018).

Hansen, K., *et al.* Genome editing with CompoZr custom zinc finger nucleases (ZFNs). *J Vis Exp* **64**, 3304 (2012).

Harlan, F.K. *et al.* Fluorogenic Substrates for Visualizing Acidic Organelle Enzyme Activities. *PLoS One* **11**, e0156312 (2016).

Hashimoto, M., Rockenstein, E., Crews, L. & Masliah, E. Role of protein aggregation in mitochondrial dysfunction and neurodegeneration in Alzheimer's and Parkinson's diseases. *Neuromolecular Med* **4**, 21-36 (2003).

Hess, G.T. *et al.* Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nat Methods* **13**, 1036-1042 (2016).

Heyer, W.D., Ehmsen, K.T. & Liu, J. Regulation of homologous recombination in eukaryotes. *Annu Rev Genet* **44**, 113-139 (2010).

Hindson, B.J. *et al.* High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal Chem* **83**, 8604-8610 (2011).

Hinnen, A., Hicks, J.B. & Fink, G.R. Transformation of yeast. *Proc Natl Acad Sci U S A* **75**, 1929-1933 (1978).

Hong, Y.B., Kim, E.Y., Yoo, H.W. & Jung, S.C. Feasibility of gene therapy in Gaucher disease using an adeno-associated virus vector. *J Hum Genet* **49**, 536-543 (2004).

Horiguchi, I., Urabe, Y., Kimura, K. & Sakai, Y. Effects of glucose, lactate and basic FGF as limiting factors on the expansion of human induced pluripotent stem cells. *J Biosci Bioeng* **125**, 111-115 (2018).

Horlbeck, M.A. *et al.* Nucleosomes impede Cas9 access to DNA in vivo and in vitro. *Elife* **5** (2016).

Horowitz, M. *et al.* Prevalence of nine mutations among Jewish and non-Jewish Gaucher disease patients. *Am J Hum Genet* **53**, 921-930 (1993).

Hsu, P.D., Lander, E.S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262-1278 (2014).

Hsu, P.D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31**, 827-832 (2013).

Hu, P. *et al.* An enrichment method to increase cell-free fetal DNA fraction and significantly reduce false negatives and test failures for non-invasive prenatal screening: a feasibility study. *J Transl Med* **17**, 124 (2019).

Huang, J.Y. *et al.* CRISPR-Cas9 generated Pompe knock-in murine model exhibits early-onset hypertrophic cardiomyopathy and skeletal muscle weakness. *Sci Rep* **10**, 10321 (2020).

Huang, W. *et al.* An induced pluripotent stem cell line (TRNDi006-A) from a MPS IIIB patient carrying homozygous mutation of p.Glu153Lys in the NAGLU gene. *Stem Cell Res* **37**, 101427 (2019).

Huang, H. *et al.* Human Pompe disease-induced pluripotent stem cells for pathogenesis modeling, drug testing and disease marker identification. *Hum Molec Genet* **20**, 4851-4864 (2011).

Huangfu, D. *et al.* Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. *Nat Biotechnol* **26**, 795-797 (2008).

Iaselli, F. *et al.* Adult-onset pulmonary involvement in Niemann-Pick disease type B. *Monaldi Arch Chest Dis* **75**, 235-240 (2011).

Isaac, R.S. *et al.* Nucleosome breathing and remodeling constrain CRISPR-Cas9 function. *Elife* **5** (2016).

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakata, A. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J Bacteriol* **169**, 5429-5433 (1987).

Itokawa, K., Tamura, N., Kawai, N., Shimazu, K. & Ishii, K. Parkinsonism in type I Gaucher's disease. *Intern Med* **45**, 1165-1167 (2006).

Jin, J. *et al.* An improved strategy for CRISPR/Cas9 gene knockout and subsequent wildtype and mutant gene rescue. *PLoS One* **15**, e0228910 (2020).

Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821 (2012).

Joung, J.K. & Sander, J.D. TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol* **14**, 49-55 (2013).

Kang, J.G., Park, J.S., Ko, J.H. & Kim, Y.S. Regulation of gene expression by altered promoter methylation using a CRISPR/Cas9-mediated epigenetic editing system. *Sci Rep* **9**, 11960 (2019).

Kaplan, A., Achord, D.T. & Sly, W.S. Phosphohexosyl components of a lysosomal enzyme are recognized by pinocytosis receptors on human fibroblasts. *Proc Natl Acad Sci U S A* **74**, 2026-2030 (1977).

Kato, R. *et al.* Parametric analysis of colony morphology of non-labelled live human pluripotent stem cells for cell quality control. *Sci Rep* **6**, 34009 (2016).

Keilani, S. *et al.* Lysosomal dysfunction in a mouse model of Sandhoff disease leads to accumulation of ganglioside-bound amyloid-β peptide. *J Neurosci* **32**, 5223-5236 (2012).

Khaddour, K. *et al.* Hematopoietic stem cell transplantation. StatPearls. TreasureIsland. (2020).

Khorrami, M., Mahdavi, M., Fakhr, F. & Kheirollahi, M. A Novel Pathogenic Variant in NAGLU (N-Acetyl-Alpha-Glucosaminidase) gene Identified by Targeted Next-Generation Sequencing Followed by in Silico Analysis. *Iran J Child Neurol* **13**, 173-183 (2019).

Kim, D. *et al.* Generation of human induced pluripotent stem cells by direct delivery of reprogramming proteins. *Cell Stem Cell* **4**, 472-476 (2009).

Kim, J.H., Chi, Y.H., Kim, G.H., Yoo, H.W. & Lee, J.H. Long-term clinical course of a patient with mucopolysaccharidosis type IIIB. *Korean J Pediatr* **59**, S37-S40 (2016).

Kim, J.M., Kim, D., Kim, S. & Kim, J.S. Genotyping with CRISPR-Cas-derived RNA-guided endonucleases. *Nat Commun* **5**, 3157 (2014).

Kim, S.U. & de Vellis, J. Stem cell-based cell therapy in neurological diseases: a review. *J Neurosci Res* **87**, 2183-2200 (2009).

Kim, Y.G., Cha, J. & Chandrasegaran, S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci U S A* **93**, 1156-1160 (1996).

Koblan, L.W. *et al.* Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat Biotechnol* (2018).

Kogut, I. *et al.* High-efficiency RNA-based reprogramming of human primary fibroblasts. *Nat Commun* **9**, 745 (2018).

Kolodny, E.H. *et al.* Phenotypic manifestations of Gaucher disease: clinical features in 48 biochemically verified type 1 patients and comment on type 2 patients. *Prog Clin Biol Res* **95**, 33-65 (1982).

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

Kong, C.W. *et al.* Risk factors for procedure-related fetal losses after mid-trimester genetic amniocentesis. *Prenat Diagn* **26**, 925-930 (2006).

Kozhich, O.A., Hamilton, R.S. & Mallon, B.S. Standardized generation and differentiation of neural precursor cells from human pluripotent stem cells. *Stem Cell Rev Rep* **9**, 531-536 (2013).

Kraushaar, D.C., Dalton, S. & Wang, L. Heparan sulfate: a key regulator of embryonic stem cell fate. *Biol Chem* **394**, 741-751 (2013).

Kwart, D., Paquet, D., Teo, S. & Tessier-Lavigne, M. Precise and efficient scarless genome editing in stem cells using CORRECT. *Nat Protoc* **12**, 329-354 (2017).

Labun, K., Montague, T.G., Gagnon, J.A., Thyme, S.B. & Valen, E. CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. *Nucleic Acids Res* **44**, W272-276 (2016).

Labun, K. *et al.* CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res* **47**, W171-W174 (2019).

Landels, E.C., Ellis, I.H., Fensom, A.H., Green, P.M. & Bobrow, M. Frequency of the Tay-Sachs disease splice and insertion mutations in the UK Ashkenazi Jewish population. *J Med Genet* **28**, 177-180 (1991).

Laoharawee, K. *et al.* Dose-Dependent Prevention of Metabolic and Neurologic Disease in Murine MPS II by ZFN-Mediated In Vivo Genome Editing. *Mol Ther* **26**, 1127-1136 (2018).

Larson, M.H. *et al.* CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat Protoc* **8**, 2180-2196 (2013).

Latour, Y.L. *et al.* Human. *Mol Genet Metab Rep* **21**, 100513 (2019).

Lau, K.T. *et al.* [Outcome of 1,355 consecutive transabdominal chorionic villus samplings in 1,351 patients]. *Chin Med J (Engl)* **118**, 1675-1681 (2005).

Laufer, B.I. & Singh, S.M. Strategies for precision modulation of gene expression by epigenome editing: an overview. *Epigenetics Chromatin* **8**, 34 (2015).

Ledford, H. & Callaway, E. Pioneers of revolutionary CRISPR gene editing win chemistry Nobel. *Nature* **586**, 346-347 (2020).

Lee, C.L. *et al.* Functional independence of Taiwanese patients with mucopolysaccharidoses. *Mol Genet Genomic Med* **7**, e790 (2019).

Lee, J.S. *et al.* Targeted PMP22 TATA-box editing by CRISPR/Cas9 reduces demyelinating neuropathy of Charcot-Marie-Tooth disease type 1A in mice. *Nucleic Acids Res* **48**, 130-140 (2020).

Leinekugel, P., Michel, S., Conzelmann, E. & Sandhoff, K. Quantitative correlation between the residual activity of beta-hexosaminidase A and arylsulfatase A and the severity of the resulting lysosomal storage disease. *Hum Genet* **88**, 513-523 (1992).

Lemonnier, T. *et al.* Modeling neuronal defects associated with a lysosomal disorder using patient-derived induced pluripotent stem cells. *Hum Mol Genet* **20**, 3653-3666 (2011).

Levy, J.M. *et al.* Cytosine and adenine base editing of the brain, liver, retina, heart and skeletal muscle of mice via adeno-associated viruses. *Nat Biomed Eng* **4**, 97-110 (2020).

Li, K., Wang, G., Andersen, T., Zhou, P. & Pu, W.T. Optimization of genome engineering approaches with the CRISPR/Cas9 system. *PLoS One* **9**, e105779 (2014).

Lin, H.Y. *et al.* Mucopolysaccharidosis III in Taiwan: Natural history, clinical and molecular characteristics of 28 patients diagnosed during a 21-year period. *Am J Med Genet A* **176**, 1799-1809 (2018).

Lin, S., Staahl, B.T., Alla, R.K. & Doudna, J.A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife* **3**, e04766 (2014).

Lin, X. Functions of heparan sulfate proteoglycans in cell signaling during development. *Development* **131**, 6009-6021 (2004).

Lino, C.A., Harper, J.C., Carney, J.P. & Timlin, J.A. Delivering CRISPR: a review of the challenges and approaches. *Drug Deliv* **25**, 1234-1257 (2018).

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

Liu, Y. *et al.* Real time PCR using TaqMan and SYBR Green for detection of Enterobacter sakazakii in infant formula. *J Microbiol Methods* **65**, 21-31 (2006).

Lu, I.L. *et al.* Correction/mutation of acid alpha-D-glucosidase gene by modified single-stranded oligonucleotides: in vitro and in vivo studies. *Gene Ther* **10**, 1910-1916 (2003).

Luciani, M., Gritti, A. & Meneghini, V. Human iPSC-Based Models for the Development of Therapeutics Targeting Neurodegenerative Lysosomal Storage Diseases. *Front Mol Biosci* **7**, 224 (2020).

Lysosomal Disease Network. "Editorial practices." Available online: https://worldsymposia.org/about-worldsymposia-lysosomal-cell-biology/. Accessed 18 January 2019.

Ma, H. *et al.* Correction of a pathogenic gene mutation in human embryos. *Nature* **548**, 413-419 (2017).

MacArthur, C.C. *et al.* Generation of human-induced pluripotent stem cells by a nonintegrating RNA Sendai virus vector in feeder-free or xeno-free conditions. *Stem Cells Int* **2012**, 564612 (2012).

Machaczka, M. *et al.* Substrate reduction therapy with miglustat for type 1 Gaucher disease: a retrospective analysis from a single institution. *Ups J Med Sci* **117**, 28-34 (2012).

Machaczka, M., Rucińska, M. & Skotnicki, A.B. [Nonmyeloablative allogeneic hematopoietic stem cell transplantation: minitransplantation]. *Przegl Lek* **56**, 633-637 (1999).

Mahmoud, H.K. *et al.* Allogeneic hematopoietic stem cell transplantation for non-malignant hematological disorders. *J Adv Res* **6**, 449-458 (2015).

Málaga, D.R. *et al.* Sensitivity, advantages, limitations, and clinical utility of targeted next-generation sequencing panels for the diagnosis of selected lysosomal storage disorders. *Genet Mol Biol* **42**, 197-206 (2019).

Mali, P., Esvelt, K.M. & Church, G.M. Cas9 as a versatile tool for engineering biology. *Nat Methods* **10**, 957-963 (2013).

Malik, N. & Rao, M.S. A review of the methods for human iPSC derivation. *Methods Mol Biol* **997**, 23-33 (2013).

Mancini, G.M., Beerens, C.E., Aula, P.P. & Verheijen, F.W. Sialic acid storage diseases. A multiple lysosomal transport defect for acidic monosaccharides. *J Clin Invest* **87**, 1329-1335 (1991).

Manning-Boğ, A.B., Schüle, B. & Langston, J.W. Alpha-synuclein-glucocerebrosidase interactions in pharmacological Gaucher models: a biological link between Gaucher disease and parkinsonism. *Neurotoxicology* **30**, 1127-1132 (2009).

Martínez-Arias, R. *et al.* Sequence variability of a human pseudogene. *Genome Res* **11**, 1071-1085 (2001).

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

Maurisse, R. *et al.* Comparative transfection of DNA into primary and transformed mammalian cells from different lineages. *BMC Biotechnol* **10**, 9 (2010).

Mazzulli, J.R. *et al.* Gaucher disease glucocerebrosidase and α-synuclein form a bidirectional pathogenic loop in synucleinopathies. *Cell* **146**, 37-52 (2011).

McCulloch, E.A. & Till, J.E. The radiation sensitivity of normal mouse bone marrow cells, determined by quantitative marrow transplantation into irradiated mice. *Radiat Res* **13**, 115-125 (1960).

Mehta, A. Epidemiology and natural history of Gaucher's disease. *Eur J Intern Med* **17 Suppl**, S2-5 (2006).

Meijer, O.L.M. *et al.* Processing of mutant N-acetyl-α-glucosaminidase in mucopolysaccharidosis type IIIB fibroblasts cultured at low temperature. *Mol Genet Metab* **122**, 100-106 (2017).

Meijer, O.L. *et al.* Residual N-acetyl-α-glucosaminidase activity in fibroblasts correlates with disease severity in patients with mucopolysaccharidosis type IIIB. *J Inherit Metab Dis* **39**, 437-445 (2016).

Meikle, P.J., Hopwood, J.J., Clague, A.E. & Carey, W.F. Prevalence of lysosomal storage disorders. *JAMA* **281**, 249-254 (1999).

Meneghini, V. *et al.* Generation of Human Induced Pluripotent Stem Cell-Derived Bona Fide Neural Stem Cells for Ex Vivo Gene Therapy of Metachromatic Leukodystrophy. *Stem Cells Transl Med* **6**, 352-368 (2017).

Mielcarek, M. *et al.* Graft-versus-host disease after nonmyeloablative versus conventional hematopoietic stem cell transplantation. *Blood* **102**, 756-762 (2003).

Mitsui, J. *et al.* Variants associated with Gaucher disease in multiple system atrophy. *Ann Clin Transl Neurol* **2**, 417-426 (2015).

Mohammed, E. E. A. & Fateen, E. M. Identification of three novel homozygous *NAGLU* mutations in Egyptian patients with Sanfilippo Syndrome B. **21** (2019).

Montague, T.G., Cruz, J.M., Gagnon, J.A., Church, G.M. & Valen, E. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res* **42**, W401-407 (2014).

Mossahebi-Mohammadi, M., Quan, M., Zhang, J.S. & Li, X. FGF Signaling Pathway: A Key Regulator of Stem Cell Pluripotency. *Front Cell Dev Biol* **8**, 79 (2020).

Mussolino, C. *et al.* A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res* **39**, 9283-9293 (2011).

Nadig, R.R. Stem cell therapy - Hype or hope? A review. *J Conserv Dent* **12**, 131-138 (2009).

Narita, A. *et al.* Ambroxol chaperone therapy for neuronopathic Gaucher disease: A pilot study. *Ann Clin Transl Neurol* **3**, 200-215 (2016).

National Institute of Neurological Disorders and Stroke. https://www.ninds.nih.gov/Disorders/Patient-Caregiver-Education/Fact-Sheets/Mucopolysaccharidoses-Fact-Sheet. Accessed 10 September 2020.

Naureckiene, S. *et al.* Identification of HE1 as the second gene of Niemann-Pick C disease. *Science* **290**, 2298-2301 (2000).

Nawrocka, D., Krzyscik, M.A., Opaliński, Ł., Zakrzewska, M. & Otlewski, J. Stable Fibroblast Growth Factor 2 Dimers with High Pro-Survival and Mitogenic Potential. *Int J Mol Sci* **21** (2020).

Neudorfer, O. *et al.* Occurrence of Parkinson's syndrome in type I Gaucher disease. *QJM* **89**, 691-694 (1996).

Norton, M.E. *et al.* Cell-free DNA analysis for noninvasive examination of trisomy. *N Engl J Med* **372**, 1589-1597 (2015).

Ojima, T. *et al.* Glycolipid dynamics in generation and differentiation of induced pluripotent stem cells. *Sci Rep* **5**, 14988 (2015).

Okita, K. *et al.* A more efficient method to generate integration-free human iPS cells. *Nat Methods* **8**, 409-412 (2011).

Ornitz, D.M. & Itoh, N. The Fibroblast Growth Factor signaling pathway. *Wiley Interdiscip Rev Dev Biol* **4**, 215-266 (2015).

Orr-Weaver, T.L., Szostak, J.W. & Rothstein, R.J. Yeast transformation: a model system for the study of recombination. *Proc Natl Acad Sci U S A* **78**, 6354-6358 (1981).

Ou, L. *et al.* ZFN-Mediated In Vivo Genome Editing Corrects Murine Hurler Syndrome. *Mol Ther* **27**, 178-187 (2019).

Ou, L. *et al.* A novel gene editing system to treat both Tay-Sachs and Sandhoff diseases. *Gene Ther* **27**, 226-236 (2020).

Oussoren, E. *et al.* Residual α-L-iduronidase activity in fibroblasts of mild to severe Mucopolysaccharidosis type I patients. *Mol Genet Metab* **109**, 377-381 (2013).

Pan, C. *et al.* Functional abnormalities of heparan sulfate in mucopolysaccharidosis-I are associated with defective biologic activity of FGF-2 on human multipotent progenitor cells. *Blood* **106**, 1956-1964 (2005).

Panicker, L.M. *et al.* Induced pluripotent stem cell model recapitulates pathologic hallmarks of Gaucher disease. *Proc Natl Acad Sci U S A* **109**, 18054-18059 (2012).

Pardo, B., Gómez-González, B. & Aguilera, A. DNA repair in mammalian cells: DNA double-strand break repair: how to fix a broken relationship. *Cell Mol Life Sci* **66**, 1039-1056 (2009).

Parenti, G., Andria, G. & Valenzano, K.J. Pharmacological Chaperone Therapy: Preclinical Development, Clinical Translation, and Prospects for the Treatment of Lysosomal Storage Disorders. *Mol Ther* **23**, 1138-1148 (2015).

Parini, R. *et al.* Open issues in Mucopolysaccharidosis type I-Hurler. *Orphanet J Rare Dis* **12**, 112 (2017).

Park, I.H. *et al.* Disease-specific induced pluripotent stem cells. *Cell* **134**, 877-886 (2008).

Pastwa, E. & Błasiak, J. Non-homologous DNA end joining. *Acta Biochim Pol* **50**, 891-908 (2003).

Patterson, M.C., Vecchio, D., Prady, H., Abel, L. & Wraith, J.E. Miglustat for treatment of Niemann-Pick C disease: a randomised controlled study. *Lancet Neurol* **6**, 765-772 (2007).

Pavan, E. *et al.* CRISPR/Cas9 Editing for Gaucher Disease Modelling. *Int J Mol Sci* **21** (2020).

Peters, C. *et al.* Hurler syndrome: II. Outcome of HLA-genotypically identical sibling and HLA-haploidentical related donor bone marrow transplantation in fifty-four children. The Storage Disease Collaborative Study Group. *Blood* **91**, 2601-2608 (1998).

Piotrowska E, Jakobkiewicz-Banecka J, Tylki-Szymanska A, Czartoryska B, Wegrzyn A, Wegrzyn G. Correlation between severity of mucopolysaccharidoses and combination of the residual enzyme activity and efficiency of GAG synthesis. *Acta Pædiatrica* **98**, *4* (2009).

Platt, F.M. Sphingolipid lysosomal storage disorders. *Nature* **510**, 68-75 (2014).

Platt, F.M. & Lachmann, R.H. Treating lysosomal storage disorders: current practice and future prospects. *Biochim Biophys Acta* **1793**, 737-745 (2009).

Poletto, E., Baldo, G. & Gomez-Ospina, N. Genome Editing for Mucopolysaccharidoses. *Int J Mol Sci* **21** (2020).

Porteus, M.H. & Carroll, D. Gene targeting using zinc finger nucleases. *Nat Biotechnol* **23**, 967-973 (2005).

Poswar, F.O. *et al.* Lysosomal diseases: Overview on current diagnosis and treatment. *Genet Mol Biol* **42**, 165-177 (2019).

Przybilla, M.J. *et al.* Comprehensive behavioral and biochemical outcomes of novel murine models of GM1-gangliosidosis and Morquio syndrome type B. *Mol Genet Metab* **126**, 139-150 (2019).

Punta, M. & Ofran, Y. The rough guide to in silico function prediction, or how to use sequence and structure information to predict protein function. *PLoS Comput Biol* **4**, e1000160 (2008).

Quarto, N. & Amalric, F. Heparan sulfate proteoglycans as transducers of FGF-2 signalling. *J Cell Sci* **107 (Pt 11)**, 3201-3212 (1994).

Rasband, W.S. Image J, U.S. National Institutes of Health, Bethesda, Maryland, USA, https://imagej.nih.gov/ij/, 1997-2008

Rath, D., Amlinger, L., Rath, A. & Lundgren, M. The CRISPR-Cas immune system: biology, mechanisms and applications. *Biochimie* **117**, 119-128 (2015).

Rees, H.A. & Liu, D.R. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat Rev Genet* **19**, 770-788 (2018).

Richardson, C.D., Ray, G.J., DeWitt, M.A., Curie, G.L. & Corn, J.E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat Biotechnol* **34**, 339-344 (2016).

Rocha, E.M. *et al.* Glucocerebrosidase gene therapy prevents α-synucleinopathy of midbrain dopamine neurons. *Neurobiol Dis* **82**, 495-503 (2015).

Rodríguez-Traver, E. *et al.* A collection of integration-free iPSCs derived from Parkinson's disease patients carrying mutations in the GBA1 gene. *Stem Cell Res* **38**, 101482 (2019).

Roitbak, T., Li, L. & Cunningham, L.A. Neural stem/progenitor cells promote endothelial cell morphogenesis and protect endothelial cells against ischemia via HIF-1alpha-regulated VEGF signaling. *J Cereb Blood Flow Metab* **28**, 1530-1542 (2008).

Romero, R. *et al.* Mechanism of glucocerebrosidase activation and dysfunction in Gaucher disease unraveled by molecular dynamics and deep learning. *Proc Natl Acad Sci U S A* **116**, 5086-5095 (2019).

Ron, I., Rapaport, D. & Horowitz, M. Interaction between parkin and mutant glucocerebrosidase variants: a possible link between Parkinson disease and Gaucher disease. *Hum Mol Genet* **19**, 3771-3781 (2010).

Roshan Lal, T. & Sidransky, E. The Spectrum of Neurological Manifestations Associated with Gaucher Disease. *Diseases* **5** (2017).

Sakuma, T., Nishikawa, A., Kume, S., Chayama, K. & Yamamoto, T. Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system. *Sci Rep* **4**, 5400 (2014).

Saleem, T. & Persaud, B. Another look at the safety effects of horizontal curvature on rural two-lane highways. *Accid Anal Prev* **106**, 149-159 (2017).

Salvioli, R. *et al.* The N370S (Asn370-->Ser) mutation affects the capacity of glucosylceramidase to interact with anionic phospholipid-containing membranes and saposin C. *Biochem J* **390**, 95-103 (2005).

Sands, M.S. & Davidson, B.L. Gene therapy for lysosomal storage diseases. *Mol Ther* **13**, 839-849 (2006).

Sanfilippo, S. *et al.* Mental retardation associated with acid mucopolysacchariduria (haparitin sulfate type 2). *J Peads*. **63**, *4* (1963).

Saraconi, G., Severi, F., Sala, C., Mattiuz, G. & Conticello, S.G. The RNA editing enzyme APOBEC1 induces somatic mutations and a compatible mutational signature is present in esophageal adenocarcinomas. *Genome Biol* **15**, 417 (2014).

Sardi, S.P. *et al.* CNS expression of glucocerebrosidase corrects alpha-synuclein pathology and memory in a mouse model of Gaucher-related synucleinopathy. *Proc Natl Acad Sci U S A* **108**, 12101-12106 (2011).

Sasaki, T. *et al.* Purification and partial characterization of alpha-N-acetylglucosaminidase from human liver. *J Biochem* **110**, 842-846 (1991).

Sasano, H. *et al.* Aromatase and 17 beta-hydroxysteroid dehydrogenase type 1 in human breast carcinoma. *J Clin Endocrinol Metab* **81**, 4042-4046 (1996).

Sasisekharan, R., Shriver, Z., Venkataraman, G. & Narayanasami, U. Roles of heparan-sulphate glycosaminoglycans in cancer. *Nat Rev Cancer* **2**, 521-528 (2002).

Savic, N. *et al.* Covalent linkage of the DNA repair template to the CRISPR-Cas9 nuclease enhances homology-directed repair. *Elife* **7** (2018).

Scharenberg, S.G. *et al.* Engineering monocyte/macrophage-specific glucocerebrosidase expression in human hematopoietic stem cells using genome editing. *Nat Commun* **11**, 3327 (2020).

Schiffmann, R. *et al.* Randomized, controlled trial of miglustat in Gaucher's disease type 3. *Ann Neurol* **64**, 514-522 (2008).

Schlaeger, T.M. *et al.* A comparison of non-integrating reprogramming methods. *Nat Biotechnol* **33**, 58-63 (2015).

Schmidtchen, A. *et al.* NAGLU mutations underlying Sanfilippo syndrome type B. *Am J Hum Genet* **62**, 64-69 (1998).

Schuchman, E.H. & Desnick, R.J. Types A and B Niemann-Pick disease. *Mol Genet Metab* **120**, 27-33 (2017).

Schuchman, E.H., Levran, O., Pereira, L.V. & Desnick, R.J. Structural organization and complete nucleotide sequence of the gene encoding human acid sphingomyelinase (SMPD1). *Genomics* **12**, 197-205 (1992).

Schueler, U.H. *et al.* Correlation between enzyme activity and substrate storage in a cell culture model system for Gaucher disease. *J Inherit Metab Dis* **27**, 649-658 (2004).

Schuh, R.S. *et al.* In vivo genome editing of mucopolysaccharidosis I mice using the CRISPR/Cas9 system. *J Control Release* **288**, 23-33 (2018).

Seki, T. *et al.* Generation of induced pluripotent stem cells from human terminally differentiated circulating T cells. *Cell Stem Cell* **7**, 11-14 (2010).

Selmer, K.K. *et al.* A mild form of Mucopolysaccharidosis IIIB diagnosed with targeted next-generation sequencing of linked genomic regions. *Eur J Hum Genet* **20**, 58-63 (2012).

Sentmanat, M.F., Peters, S.T., Florian, C.P., Connelly, J.P. & Pruett-Miller, S.M. A Survey of Validation Strategies for CRISPR-Cas9 Editing. *Sci Rep* **8**, 888 (2018).

Shapiro, E.G., Lockman, L.A., Balthazor, M. & Krivit, W. Neuropsychological outcomes of several storage diseases with and without bone marrow transplantation. *J Inherit Metab Dis* **18**, 413-429 (1995).

Shapiro, E. *et al.* Analysis of the caregiver burden associated with Sanfilippo syndrome type B: panel recommendations based on qualitative and quantitative data. *Orphanet J Rare Dis* **14**, 168 (2019).

Sharma, A. & Changotra, H. Mutagenic primer-based PCR-RFLP assay for genotyping IRGM gene promoter variant rs4958843 (C/T). *J Clin Lab Anal* **32**, e22346 (2018).

Shihabuddin, L.S. *et al.* Intracerebral transplantation of adult mouse neural progenitor cells into the Niemann-Pick-A mouse leads to a marked decrease in lysosomal storage pathology. *J Neurosci* **24**, 10642-10651 (2004).

Sidransky, E. *et al.* Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. *N Engl J Med* **361**, 1651-1661 (2009).

Singh, S.K., Kumar, R. & Wengel, J. Synthesis of Novel Bicyclo[2.2.1] Ribonucleosides: 2'-Amino- and 2'-Thio-LNA Monomeric Nucleosides. *J Org Chem* **63**, 6078-6079 (1998).

Singh, U. *et al.* Novel live alkaline phosphatase substrate for identification of pluripotent stem cells. *Stem Cell Rev Rep* **8**, 1021-1029 (2012).

Sivakumur, P. & Wraith, J.E. Bone marrow transplantation in mucopolysaccharidosis type IIIA: a comparison of an early treated patient with his untreated sibling. *J Inherit Metab Dis* **22**, 849-850 (1999).

Smid, B.E. *et al.* Biochemical response to substrate reduction therapy versus enzyme replacement therapy in Gaucher disease type 1 patients. *Orphanet J Rare Dis* **11**, 28 (2016).

Sohn, Y.B. *et al.* Phase I/II clinical trial of enzyme replacement therapy with idursulfase beta in patients with mucopolysaccharidosis II (Hunter syndrome). *Orphanet J Rare Dis* **8**, 42 (2013).

Soko, N.D., Masimirembwa, C. & Dandara, C. A cost effective RFLP method to genotype Solute carrier organic anion 1B1 (SLCO1B1) c.1929A>C (p.Leu643Phe, rs34671512); a variant with potential effect on rosuvastatin pharmacokinetics. *BMC Res Notes* **11**, 384 (2018).

Somaraju, U.R. & Tadepalli, K. Hematopoietic stem cell transplantation for Gaucher disease. *Cochrane Database Syst Rev* **10**, CD006974 (2017).

Son, M.Y., Kim, H.J., Kim, M.J. & Cho, Y.S. Physical passaging of embryoid bodies generated from human pluripotent stem cells. *PLoS One* **6**, e19134 (2011).

Song, H.Y. *et al.* Using CRISPR/Cas9-Mediated GLA Gene Knockout as an In Vitro Drug Screening Model for Fabry Disease. *Int J Mol Sci* **17** (2016).

Spillantini, M.G. *et al.* Alpha-synuclein in Lewy bodies. *Nature* **388**, 839-840 (1997).

Squeo, G.M. *et al.* Customised next-generation sequencing multigene panel to screen a large cohort of individuals with chromatin-related disorder. *J Med Genet* **57**, 760-768 (2020).

Statistics Canada. Canadian Community Health Survey, Statistics Canada, 3226 (2001).

Štefková, K., Procházková, J. & Pacherník, J. Alkaline phosphatase in stem cells. *Stem Cells Int* **2015**, 628368 (2015).

Stirnemann, J. *et al.* A Review of Gaucher Disease Pathophysiology, Clinical Presentation and Treatments. *Int J Mol Sci* **18** (2017).

Sterner, E. *et al.* FGF-FGFR signaling mediated through glycosaminoglycans in microtiter plate and cell-based microarray platforms. *Biochem* **52**, 9009-9019 (2013).

Stone, D.L. & Sidransky, E. Hydrops fetalis: lysosomal storage disorders in extremis. *Adv Pediatr* **46**, 409-440 (1999).

Sullivan, S. *et al.* Quality control guidelines for clinical-grade human induced pluripotent stem cell lines. *Regen Med* **13**, 859-866 (2018).

Sun, Y. *et al.* Properties of neurons derived from induced pluripotent stem cells of Gaucher disease type 2 patient fibroblasts: potential role in neuropathology. *PLoS One* **10**, e0118771 (2015).

Sürün, D. *et al.* Efficient Generation and Correction of Mutations in Human iPS Cells Utilizing mRNAs of CRISPR Base Editors and Prime Editors. *Genes (Basel)* **11** (2020).

Suzuki, K. *et al.* Neuronal accumulation of alpha- and beta-synucleins in the brain of a GM2 gangliosidosis mouse model. *Neuroreport* **14**, 551-554 (2003).

Swan, M. & Saunders-Pullman, R. The association between ß-glucocerebrosidase mutations and parkinsonism. *Curr Neurol Neurosci Rep* **13**, 368 (2013).

Taapken, S.M. *et al.* Karotypic abnormalities in human induced pluripotent stem cells and embryonic stem cells. *Nat Biotechnol* **29**, 313-314 (2011).

Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-676 (2006).

Tamargo, R.J., Velayati, A., Goldin, E. & Sidransky, E. The role of saposin C in Gaucher disease. *Mol Genet Metab* **106**, 257-263 (2012).

Tan, S. *et al.* Non-viral vector based gene transfection with human induced pluripotent stem cells derived cardiomyocytes. *Sci Rep* **9**, 14404 (2019).

Tarailo-Graovac, M. *et al.* Identification of a large intronic transposal insertion in *SLC17A5* causing sialic acid storage disease. *Orphanet J Rare Dis* **12**, 28 (2017).

Tayebi, N. *et al.* Gaucher disease and parkinsonism: a phenotypic and genotypic characterization. *Mol Genet Metab* **73**, 313-321 (2001).

Taylor K.R., Rudisill, J.A., & Gallo, R.L. Structural and sequence motifs in dermatan sulfate for promoting fibroblast growth factor-2 (FGF-2) and FGF-7 activity. *J Biol Chem* **280**, 5300-5306 (2005).

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

Tofoli, F.A., Chien, H.F., Barbosa, E.R. & Pereira, L.V. Generation of three human induced pluripotent stem cell (hiPSC) lines derived from one Gaucher disease patient with Parkinson's disease and two unrelated Parkinson's disease patients with GBA mutations. *Stem Cell Res* **39**, 101519 (2019).

Tseng, L.H. *et al.* Simultaneous genotyping of single nucleotide polymorphisms in the IL-6, IL-10, TNFalpha and TNFbeta genes. *Tissue Antigens* **59**, 280-286 (2002).

Tsui, N.B. *et al.* Noninvasive prenatal diagnosis of hemophilia by microfluidics digital PCR analysis of maternal plasma DNA. *Blood* **117**, 3684-3691 (2011).

Tucker, B.A., Anfinson, K.R., Mullins, R.F., Stone, E.M. & Young, M.J. Use of a synthetic xeno-free culture substrate for induced pluripotent stem cell induction and retinal differentiation. *Stem Cells Transl Med* **2**, 16-24 (2013).

Turner, K. Reprogramming skin fibroblasts from patients with Gaucher disease to induced pluripotent stem cells for *in vitro* disease modeling. Hons. Thesis. University of Victoria. (2015).

Tylki-Szymańska, A. *et al.* Non-neuronopathic Gaucher disease due to saposin C deficiency. *Clin Genet* **72**, 538-542 (2007).

Tyndall, A. *et al.* Treatment of systemic sclerosis with autologous haemopoietic stem cell transplantation. *Lancet* **349**, 254 (1997).

Valetdinova, K.R., Ovechkina, V.S. & Zakian, S.M. Methods for Correction of the Single-Nucleotide Substitution c.840C>T in Exon 7 of the SMN2 Gene. *Biochemistry (Mosc)* **84**, 1074-1084 (2019).

Valstar, M.J., Ruijter, G.J., van Diggelen, O.P., Poorthuis, B.J. & Wijburg, F.A. Sanfilippo syndrome: a mini-review. *J Inherit Metab Dis* **31**, 240-252 (2008).

van Dussen, L., Biegstraaten, M., Hollak, C.E. & Dijkgraaf, M.G. Cost-effectiveness of enzyme replacement therapy for type 1 Gaucher disease. *Orphanet J Rare Dis* **9**, 51 (2014).

van Horssen, J., Wesseling, P., van den Heuvel, L.P., de Waal, R.M. & Verbeek, M.M. Heparan sulphate proteoglycans in Alzheimer's disease and amyloid-related disorders. *Lancet Neurol* **2**, 482-492 (2003).

Várkonyi, J., Simon, Z., Soós, K. & Poros, A. Gaucher disease type I complicated with Parkinson's syndrome. *Haematologia (Budap)* **32**, 271-275 (2002).

Vellodi, A., Young, E., New, M., Pot-Mees, C. & Hugh-Jones, K. Bone marrow transplantation for Sanfilippo disease type B. *J Inherit Metab Dis* **15**, 911-918 (1992).

Vitner, E.B., Futerman, A.H. & Platt, N. Innate immune responses in the brain of sphingolipid lysosomal storage diseases. *Biol Chem* **396**, 659-667 (2015).

Vogelstein, B. & Kinzler, K.W. Digital PCR. *Proc Natl Acad Sci U S A* **96**, 9236-9241 (1999).

Vouillot, L., Thélie, A. & Pollet, N. Comparison of T7E1 and surveyor mismatch cleavage assays to detect mutations triggered by engineered nucleases. *G3 (Bethesda)* **5**, 407-415 (2015).

Wan, L. *et al.* Mutation analysis of Gaucher disease patients in Taiwan: high prevalence of the RecNciI and L444P mutations. *Blood Cells Mol Dis* **36**, 422-425 (2006).

Wang, D. *et al.* Optimized CRISPR guide RNA design for two high-fidelity Cas9 variants by deep learning. *Nat Commun* **10**, 4284 (2019).

Wang, J. *et al.* Allogeneic Hematopoietic Stem Cell Transplantation in Thirty-Four Pediatric Cases of Mucopolysaccharidosis-A Ten-Year Report from the China Children Transplant Group. *Biol Blood Marrow Transplant* **22**, 2104-2108 (2016).

Weber, B. *et al.* Sanfilippo type B syndrome (mucopolysaccharidosis III B): allelic heterogeneity corresponds to the wide spectrum of clinical phenotypes. *Eur J Hum Genet* **7**, 34-44 (1999).

Weber, B., Hopwood, J.J. & Yogalingam, G. Expression and characterization of human recombinant and alpha-N-acetylglucosaminidase. *Protein Expr Purif* **21**, 251-259 (2001).

Weinreb, N.J. *et al.* Effectiveness of enzyme replacement therapy in 1028 patients with type 1 Gaucher disease after 2 to 5 years of treatment: a report from the Gaucher Registry. *Am J Med* **113**, 112-119 (2002).

Weinreb, N.J. *et al.* Life expectancy in Gaucher disease type 1. *Am J Hematol* **83**, 896-900 (2008).

Welling, L. *et al.* Early Umbilical Cord Blood-Derived Stem Cell Transplantation Does Not Prevent Neurological Deterioration in Mucopolysaccharidosis Type III. *JIMD Rep* **18**, 63-68 (2015).

Wijburg, F.A., Węgrzyn, G., Burton, B.K. & Tylki-Szymańska, A. Mucopolysaccharidosis type III (Sanfilippo syndrome) and misdiagnosis of idiopathic developmental delay, attention deficit/hyperactivity disorder or autism spectrum disorder. *Acta Paediatr* **102**, 462-470 (2013).

Winder-Rhodes, S.E. *et al.* Genetic and pathological links between Parkinson's disease and the lysosomal disorder Sanfilippo syndrome. *Mov Disord* **27**, 312-315 (2012).

Wilkinson, F.L. *et al.* Neuropathology in mouse models of mucopolysaccharidosis type I, IIIA and IIIB. *PLoS One* **7**, e35787 (2012).

Wong, N., Liu, W. & Wang, X. WU-CRISPR: characteristics of functional guide RNAs for the CRISPR/Cas9 system. *Genome Biol* **16**, 218 (2015).

Woodard, C.M. *et al.* iPSC-derived dopamine neurons reveal differences between monozygotic twins discordant for Parkinson's disease. *Cell Rep* **9**, 1173-1182 (2014).

Wu, X., Kriz, A.J. & Sharp, P.A. Target specificity of the CRISPR-Cas9 system. *Quant Biol* **2**, 59-70 (2014).

Xie, F. *et al.* Seamless gene correction of β-thalassemia mutations in patient-specific iPSCs using CRISPR/Cas9 and piggyBac. *Genome Res* **24**, 1526-1533 (2014).

Xu, H. & Ren, D. Lysosomal physiology. *Annu Rev Physiol* **77**, 57-80 (2015).

Xu, Y.H. *et al.* Accumulation and distribution of α-synuclein and ubiquitin in the CNS of Gaucher disease mouse models. *Mol Genet Metab* **102**, 436-447 (2011).

Xu, Y.H. *et al.* Multiple pathogenic proteins implicated in neuronopathic Gaucher disease mice. *Hum Mol Genet* **23**, 3943-3957 (2014).

Yamanaka, S. *et al.* Apolipoprotein B mRNA-editing protein induces hepatocellular carcinoma and dysplasia in transgenic animals. *Proc Natl Acad Sci U S A* **92**, 8483-8487 (1995).

Yang, D. *et al.* Enrichment of G2/M cell cycle phase in human pluripotent stem cells enhances HDR-mediated gene repair with customizable endonucleases. *Sci Rep* **6**, 21264 (2016).

Yeh, W.H., Chiang, H., Rees, H.A., Edge, A.S.B. & Liu, D.R. In vivo base editing of post-mitotic sensory cells. *Nat Commun* **9**, 2184 (2018).

Yiangou, L. *et al.* Method to Synchronize Cell Cycle of Human Pluripotent Stem Cells without Affecting Their Fundamental Characteristics. *Stem Cell Reports* **12**, 165-179 (2019).

Yogalingam, G. & Hopwood, J.J. Molecular genetics of mucopolysaccharidosis type IIIA and IIIB: Diagnostic, clinical, and biological implications. *Hum Mutat* **18**, 264-281 (2001).

Yogalingam, G., Weber, B., Meehan, J., Rogers, J. & Hopwood, J.J. Mucopolysaccharidosis type IIIB: characterisation and expression of wild-type and mutant recombinant alpha-N-acetylglucosaminidase and relationship with sanfilippo phenotype in an attenuated patient. *Biochim Biophys Acta* **1502**, 415-425 (2000).

Ysselstein, D., Shulman, J.M. & Krainc, D. Emerging links between pediatric lysosomal storage diseases and adult parkinsonism. *Mov Disord* **34**, 614-624 (2019).

Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917-1920 (2007).

Yu, X. *et al.* Improved delivery of Cas9 protein/gRNA complexes using lipofectamine CRISPRMAX. *Biotechnol Lett* **38**, 919-929 (2016).

Zak, B.M., Crawford, B.E. & Esko, J.D. Hereditary multiple exostoses and heparan sulfate polymerization. *Biochim Biophys Acta* **1573**, 346-355 (2002).

Zech, M. *et al.* Niemann-Pick C disease gene mutations and age-related neurodegenerative disorders. *PLoS One* **8**, e82879 (2013).

Zeevi, D.A. *et al.* Proof-of-principle rapid noninvasive prenatal diagnosis of autosomal recessive founder mutations. *J Clin Invest* **125**, 3757-3765 (2015).

Zeng, J. *et al.* Therapeutic base editing of human hematopoietic stem cells. *Nat Med* **26**, 535-541 (2020).

Zhang, H. *et al.* CRISPR/Cas9-Mediated Gene Editing in Human iPSC-Derived Macrophage Reveals Lysosomal Acid Lipase Function in Human Macrophages-Brief Report. *Arterioscler Thromb Vasc Biol* **37**, 2156-2160 (2017).

Zhao, H.G., Aronovich, E.L. & Whitley, C.B. Genotype-phenotype correspondence in Sanfilippo syndrome type B. *Am J Hum Genet* **62**, 53-63 (1998).

Zhao, H.G., Li, H.H., Bach, G., Schmidtchen, A. & Neufeld, E.F. The molecular basis of Sanfilippo syndrome type B. *Proc Natl Acad Sci U S A* **93**, 6101-6105 (1996).

Zhao, K.W. & Neufeld, E.F. Purification and characterization of recombinant human alpha-N-acetylglucosaminidase secreted by Chinese hamster ovary cells. *Protein Expr Purif* **19**, 202-211 (2000).

Zhao, L.R. *et al.* Human bone marrow stem cells exhibit neural phenotypes and ameliorate neurological deficits after grafting into the ischemic brain of rats. *Exp Neurol* **174**, 11-20 (2002).

Zhou, C. *et al.* Highly efficient base editing in human tripronuclear zygotes. *Protein Cell* **8**, 772-775 (2017).

Zimran, A., Altarescu, G. & Elstein, D. Pilot study using ambroxol as a pharmacological chaperone in type 1 Gaucher disease. *Blood Cells Mol Dis* **50**, 134-137 (2013).

Zimran, A., Gelbart, T., Westwood, B., Grabowski, G.A. & Beutler, E. High frequency of the Gaucher disease mutation at nucleotide 1226 among Ashkenazi Jews. *Am J Hum Genet* **49**, 855-859 (1991).

Zischewski, J., Fischer, R. & Bortesi, L. Detection of on-target and off-target mutations generated by CRISPR/Cas9 and other sequence-specific nucleases. *Biotechnol Adv* **35**, 95-104 (2017).

Zlotogora, J., Sagi, M., Zeigler, M. & Bach, G. Gaucher disease type I and pregnancy. *Am J Med Genet* **32**, 475-477 (1989).