# Hematopoietic stem cell gene therapy for Pompe disease using a novel recombinant form of acid-alpha glucosidase

Candidate:

Slawomir Wojciech Wantuch

Supervisors:

Prof Bobby Gaspar

Dr Giuseppa Piras

Great Ormond Street

Institute of Child Health

University College London

A thesis submitted for the degree of Doctor of Philosophy

## Declaration:

I, Slawomir Wantuch confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signature………………………………….

Date……11 April 2022…………..

## Abstract:

Glycogen storage disease type II, also known as Pompe disease (PD), is caused by aberrant or absent expression of the lysosomal enzyme acid-alfa-glucosidase (GAA) that is responsible for the breakdown of glycogen into glucose within cells. Babies born with PD suffer from muscle weakness, cardiac myopathy and consequently without treatment they die in the 1st year of life due to respiratory and cardiac failure. In contrast, adolescence PD has a milder course and adverse effects. Currently, patients with PD are treated with enzyme replacement therapy (ERT), where recombinant GAA enzyme is intravenously (i.v.) injected and through bloodstream distributed into lysosomes utilising the mannose-6-phosphate receptor (M6PR). However, ERT is limited by the host humoral response to exogenous recombinant GAA, the lack of penetration the blood-brain-barrier, and finally, an inadequate expression of M6PR on muscle tissues, limiting cross-correction in skeletal muscle.

Autologous Hematopoietic Stem Cells (HSCs) Gene Therapy for PD could potentially provide constant and life-long source of endogenous GAA enzyme. Due inadequate expression of M6PRs on  skeletal muscle cells limiting cross-correction, this study aims to interrogate alternative approach : using a novel muscle homing peptides (MHPs) (X. Gao et al., 2014; Samoylova & Smith, 1999), which shown strong affinity for skeletal muscle and proving better GAA uptake into these tissues. For this purpose, MHP was incorporated into the N-terminus of GAA. In addition, a synthetic secretion peptide (SP) was placed upstream of MHP to improve secretion of the enzyme from producing cells.

*In vitro* study showed advantageous uptake of MHP-GAA protein to *Gaa -/-* murine myoblast. The addition of the SP led to 2.5-fold increase of GAA enzymes being secreted from transduced cells *in vitro*. This project also proposed a cellular receptor involved in the interaction with the MHP.

Furthermore, an *in vivo* evaluation of cross-correction in Pompe mouse model was conducted confirming long-term GAA transgene expression in transduced and transplanted HSCs. Also, cross-correction and reduction of glycogen in harvested tissues

in GT treated *Gaa -/-* mice was observed, further supporting GT as a potential treatment to alleviate Pompe disease symptoms.

## Impact Statement:

In recent years cell and gene therapy have been seen an exponential growth of clinical trials in rare and severe conditions giving hope to thousands of patients with unmet clinical need and opening the era of personalised medicine. AAV *in vivo* gene therapies were already approved for spinal muscular atrophy, inherited blindness, or haemophilia. While lentiviral *ex vivo* gene therapy was approved for the treatment of Metachromatic Leukodystrophy, a long-awaited ray of light in this filed, and several other clinical trials are in pipeline for rare metabolic inherited disorders.

This thesis addresses the development of an efficacious HSCs gene therapy for Pompe disease. It depicts steps to improve human acid alpha glucosidase (hGAA) production, secretion and uptake aiming at the cross-correction and symptoms relief of Pompe disease by inserting an engineered hGAA through a lentiviral vector into HSCs. Chapter 3.1 focuses on redesigning the native N-terminus of hGAA protein and equipping it with a novel muscle homing tag to facilitate its greater uptake into target cells. In addition, this chapter provides insights on secretory peptides to maximise hGAA protein secretion boosting gene therapy outcomes and lowering potential toxic intracellular accumulation of the protein in question. Furthermore, chapter 3.2 focuses on the transgene GAA expression under different mammalian promoters, and posttranslational modification of hGAA protein produced by transduced cell lines. Pompe like mouse myotubes were generated as model to verify the potency of a muscle homing tag in uptake, lysosomal targeting, and glycogen reduction upon exposure to rhGAA in transwell settings. Then, Chapter 3.3 shows the efficacy of HSPCs gene therapy using the herein developed rhGAA in *Gaa -/-* mouse model. This study proved the ability of transduced and transplanted HSCs to secrete active form of hGAA protein into a bloodstream, which then was able of cross-correction and partial reduction of glycogen burden in harvested tissues.

Several studies including Duchenne Muscular Dystrophy utilised muscle homing tag to deliver intracellularly therapeutic agents, however no study yet elucidated its likely cellular entry mechanism. Chapter 3.4 have potentially identified a novel interaction of muscle homing tag and its cellular counterpart. This plasma membrane receptor

identified through RNA-seq might serve as a prospective candidate for therapeutic delivery.

While this report describes applications in gene therapy field, the findings depicted here might well serve a broad range of biologic research, offering a multipurpose mean for the expression of a transgene cassette and downstream translation of recombinant proteins. Furthermore, proposed approach of remodelling of N-terminus of proteins, including addition of amino acids linkers or secretory signal peptide, might be applied in designing polypeptides in specific ways to convey desired effects in therapy/drug deliver. In addition, proposed novel interaction might well open an exciting new way to therapeutic targeting across a varied range of diseases.

## Acknowledgments:

I would like to thank my supervisors Prof Bobby Gaspar and Dr Giuseppa Piras for their patience, support, guidance, and constructive criticism, which helped me to mature and sharpened as a scientist. On my journey through science and beyond your wisdom and friendship shall guide me.

I would like to give my thanks to every member of the Gaspar group for their help and words of encouragement. In particular, I would like to extend my thanks to Ms Agnes Chan – working with you was a great privilege and fun.

Furthermore, I would like to thank to all members of the Molecular and Cellular Immunology department for their kindness and time to answer my endless scientific questions and in particular Ms Marta Zinicola and Dr Jenny Yeung.

I would like to thank Ms Patricia Cheng for continuous support and constant smile on her face.

Finally, I would like to give my special thanks to those without whom I would have neither the strength, courage nor determination to pursuit my dreams - my beloved family and my partner.

Dziękuję za Waszą wiarę we mnie oraz niekończące się słowa otuchy i wsparcia. Jesteście jednym z filarów na których opiera się mój świat.

Grazie per la tua pazienza, e il tempo dedicatomi.

# Table of Contents:

## List of Figures:

## List of Tables:

## Abbreviations:

Acid alpha-glucosidase .................................................................................**GAA**

Adeno-associated virus .................................................................................**AAV**

Adaptor protein .................................................................................. **AP2**

Adaptor protein – phosphorylated ......................................................... **AP2-P**

American Food and Drug Administration ................................................ **FDA**

Amino acid ...................................................................................... **aa**

Antisense oligonucleotides ............................................................**AONs**

Arylsulfatase A ..............................................................................**ARSA**

Aspartate aminotransferase ..........................................................**AST**

Aspartylglucosaminuria .................................................................. **AGU**

Blood-brain barrier ...................................................................... **BBB**

Burst-forming unit ......................................................................... **BFU**

Cardiac output ............................................................................ **CO**

Cation-independent mannose-6-phosphate receptor .........................................**CI-M6PR**

Cell penetrating peptide ...............................................................**CPP**

Central nervous system ................................................................ **CNS**

Central polypurine tract ................................................................ **cPPT**

Central termination sequence ........................................................**CTS**

Ceramide glucosyltransferase ..........................................................**GlcCer synthase**

Chaperone N-butyldeoxynojirimycin ............................................... **NB-DNJ**

Chemokine receptor 4 ................................................................... **CXCR4**

Clathrin-mediated endocytosis .......................................................... **CME**

Clustered regularly interspaced short palindromic repeats/caspase 9 ........**CRISPR/Cas9**

Colony forming unit ........................................................................................ **CFU**

Common lymphoid progenitors ...................................................................... **CLP**

Common myeloid progenitors ......................................................................... **CMP**

Creatine kinase ..................................................................................................**CK**

CRISPR-associated protein 9 .......................................................................... **Cas9**

Cross Reactive Immunologic Material ......................................................... **CRIM**

Dimerization signal ..........................................................................................**DIS**

Deoxyribonucleic acid ................................................................................... **DNA**

Duchenne muscular dystrophy .....................................................................**DMD**

Ejection fraction ............................................................................................. **EF**

Electrocardiogram ......................................................................................... **ECG**

Electromyogram ............................................................................................ **EMG**

Elongation factor-1α ......................................................................................**EF1α**

Endoplasmic reticulum ....................................................................................**ER**

Endothelial progenitor cells ......................................................................... **EPCs**

Enzyme replacement therapy .......................................................................**ERT**

Fluorescence-activated cell sorting ..............................................................**FACS**

Forced vital capacity .......................................................................................**FVC**

Glucose tetrasaccharide ................................................................................**Glc4**

Granulocyte ...................................................................................................... **G**

Granulocyte, erythrocyte, monocyte, megakaryocyt ............................................**GEMM**

Granulocyte–macrophage ............................................................................... **GM**

Glyceraldehyde 3-phosphate dehydrogenase ....................................................**GAPDH**

Glycogen debranching enzyme ....................................................................**GDE**

Glycogen phosphorylase ....................................................................... **GP**

Glycogen storage disease type II ...................................................................**GSDII**

Glycogen synthetase ..................................................................................**GYS**

Glycogen synthetase 1 .......................................................................... **GYS1**

Glycogenin .............................................................................................**GYG**

Glycosylation-independent lysosomal transport .......................................................**GILT**

Haematopoietic stem cell transplant .......................................................**HSCT**

Hematopoietic stem cells ........................................................................**HSCs**

Inference of CRISPR edits **.......................................................................ICE**

Immunofluorescence ............................................................................... **IF**

Immunoglobulin G ...................................................................................**IgG**

Induced pluripotent stem cells ..............................................................**iPSC**

Infant Onset Pompe Disease ...................................................................**IOPD**

Insulin-like growth factor 1 ..................................................................... **IGF-1**

Insulin-like growth factor binding protein-3 ........................................**IGFBP-3**

Insulin-like growth factor II ...................................................................... **IGF-II**

Integrated DNA Technology ...................................................................... **IDT**

Intraventricular septum diastole ..................................................................**IVS diastole**

Interventricular septum systole ................................................................... **IVS systole**

Inverted terminal repeats .........................................................................**ITRs**

*In vitro* immortalisation ................................................................................**IVIM**

Knockout ............................................................................................ **KO**

Lactate dehydrogenase ..................................................................... **LDH**

Late Onset Pompe Disease ...............................................................**LOPD**

Left ventricle mass ..............................................................................**LV**

Left ventricular posterior wall diastole ................................................**LVPW D**

Left ventricular posterior wall systole ................................................ **LVPW S**

Lentivirus ............................................................................................**LV**

Light chain and heavy chain of clathrin complex ...........................................**CLTA/CLTC**

Liquid culture ................................................................................. **LC**

Low density lipoprotein receptor................................................... **LDLR**

Lysosomal acid lipase ......................................................................**LIPA**

Lysosomal storage disorder ............................................................**LSD**

Lysosomes .......................................................................................**Lys**

Macrophage ...................................................................................... **M**

Magnetic-activated cell sorting buffer................................................... **MACS**

Major histocompatibility complex II ...................................................**MHC II**

Maltase-glucoamylase ......................................................................**MGA**

Mannose 6-phosphate receptor ................................................... **M6PR**

Mannose receptor ........................................................................ **ManR**

Mannose-6-phosphate ................................................................... **M6P**

Mass spectroscopy ..........................................................................**MS**

Mesenchymal stem cells .................................................................**MSC**

Messenger RNA .......................................................................... **mRNA**

Quantitative polymerase chain reaction ................................................................. **qPCR**

Receptor associated protein ...................................................................................... **RAP**

Recombinant adeno-associated virus ..................................................................... **rAAV**

Recombinant human acid alpha-glucosidase ........................................................ **rhGAA**

Regulatory T cells ...................................................................................................... **Treg**

Remain Reverse response element ............................................................................ **RRE**

Reverse transcription polymerase chain reaction ............................................... **RT-qPCR**

Ribonucleic acid ......................................................................................................... **RNA**

Room temperature ........................................................................................................ **RT**

Scavenger receptor B1 family ................................................................................... **SR-BI**

Scavenger receptor B1 family variant splicing BII ................................................. **SR-BII**

Self-complementary ...................................................................................................... **sc**

Signal recognition particle ........................................................................................ **SRP**

Silico secretion peptide .............................................................................................. **SSP**

Single guided RNA ................................................................................................... **sgRNA**

Single stranded DNA ................................................................................................ **ssDNA**

Six-minute walk test ................................................................................................ **6MWT**

Small nuclear RNAs .............................................................................................. **U1snRNA**

Sodium potassium ATPase pump ...................................................................... **Na⁺K⁺ATPase**

Spleen focus-forming virus promoter ......................................................................... **SF**

Spleen focus-forming virus ...................................................................................... **SFFV**

Splice acceptor ............................................................................................................. **SA**

Splice donor ................................................................................................................. **SD**

Substrate reduction therapy ...................................................................................**SRT**

The American Food and Drug Administration ........................................................ **FDA**

Tissue specific stem cells ...................................................................................**TSSCs**

Transforming growth factor β ............................................................**TGF-β-dependent**

Tricarboxylic Acid Cycle ........................................................................................**TCA**

Variance Stabilizing Transformation ....................................................................**VST**

Vector copy number .............................................................................................**VCN**

Vesicular stomatitis virus ..................................................................................... **VSV**

Wild type ............................................................................................................... **WT**

Woodchuck Hepatitis virus posttranscriptional regulatory element ...................... **WPRE**

α-galactosidase A ................................................................................................. **GLA**

β- galactosidase .................................................................................................. **β-Gal**

β-globin locus control region ...............................................................................**LCR**

β-hexosaminidase A ...........................................................................................**HEXA**

# Chapter 1   Introduction

## 1.1. Lysosomal storage disorders (LSDs)

Lysosomal storage disorders (LSDs) are a group of more than 50 rare inborn errors of metabolism (Hers, 1963; Neufeld, 1991; Vellodi, 2005), most of which are inherited as autosomal recessive traits (Platt et al., 2018). LSDs are caused either by a deficient function of lysosomal enzymes, integral membrane proteins, lysosomal activators or protein involved in lysosomal hydrolases flux (Neufeld, 1991; Winchester et al., 2000). Notably, the majority of mutations are present in genes encoding catabolic enzymes (acid hydrolases) such as proteases and glycosidases responsible for the breakdown of macromolecules (Gritti, 2011). In the absence of enzyme activity, there is an accumulation of these macromolecules over time and leading consequently to cell dysfunction, cell death and progressive clinical manifestation (Winchester et al., 2000). The majority of lysosomal enzymes are ubiquitously expressed, therefore the malfunction in a particular enzyme leads to irreversible damage to generic cellular organelles/ mechanisms which affects multi organ systems including bones, skeletal muscle, immune system, kidneys, liver, spleen, lungs, heart, central nervous system (CNS)(Gritti, 2011; Platt et al., 2018; Scriver, 1969). Particularly, symptoms presenting across LSDs comprise ischemic stroke, seizures, hepatosplenomegaly, neurodegenerative and musculoskeletal manifestations (Mueller et al., 2013; Ranieri et al., 2016; Staretz-Chacham et al., 2009; vom Dahl & Mengel, 2010). Current diagnosis of LSDs is based primarily on clinical manifestation, followed by enzymatic activity assays and gene sequencing (Parenti, Andria, & Ballabio, 2015). However, the identification of milder forms of LSD might be delayed due to the mild nature of the clinical signs causing irreversible damage to affected organs (Parenti, Andria, & Ballabio, 2015). Worldwide incidences of LSDs range from approximately 1:40,000 birth for Pompe disease (PD), 1:57,000 for Gaucher's disease, to 1:100,000 for neuronal ceroid lipofuscinoses (NCLs) (Haltia, 2006); however, as a group the LSD incidence is as high as 1:8000 live births (Fuller et al., 2006; Meikle et al., 1999). While for some LSDs a treatment is available, for the majority of them a treatment does not exist yet and they need to be managed symptomatically only (Platt et al., 2018). However, at present there are many therapies

undergoing phase I/II clinical trials like allogenic haematopoietic stem cell transplant (HSCT) (Orchard & Tolar, 2010), gene therapy utilising AAV or LV vectors (Massaro et al., 2021) , more advanced forms of enzyme replacement therapy (ERT) (Rohrbach & Clarke, 2007), substrate reduction therapy (SRT) (Coutinho et al., 2017), pharmacological chaperone therapy (PCT) (Beck, 2010b; Haneef & Doss, 2016; Smid et al., 2010) or stop-codon read-through (Dardis & Buratti, 2018), giving hope to affected individuals to alleviate their symptoms, stabilise their clinical condition and lead a normal life (Gritti, 2011).

### 1.1.1. Classification of LSDs

There are several classifications of LSDs depending on the type of mutation or clinical severity of the disease (Winchester et al., 2000). However, the most descriptive and used is a classification according to the type of deficient enzymes affected and undigested macromolecules accumulated inside lysosomes and/or cellular matrix, in which we distinguish mucopolysaccharidoses, glycoproteinoses and sphingolipidoses. Those and several classes of non-enzyme related defects (Platt et al., 2018; Sleat et al., 2009) are presented on Table 1.1

**Table 1.1 Summary of the LSDs.**

| Category | Disease | Defective protein | Main storage material |
|---|---|---|---|
| **Mucopolysaccharidoses (MPSs)** | MPS I (Hurler, Scheie) | α-Iduronidase | Heparan sulphate, dermatan sulphate, |
| | MPS II (Hunter) | Iduronate sulphatase | Heparan sulphate, dermatan sulphate, |
| | MPS III A (Sanfilippo A) | Heparan sulphamidase | Heparan sulphate |
| | MPS III B (Sanfilippo B) | Acetyl α-glucosaminidase | Heparan sulphate |
| | MPS III C (Sanfilippo C) | Acetyl CoA: α-glucosaminide, N-acetyltransferase | Heparan sulphate |
| | MPS III D (Sanfilippo D) | N-acetyl glucosamine-6-sulphatase | Heparan sulphate |
| | MPS IVA (Morquio A) | N-acetyl galactosamine-6-sulphatase | Keratan sulphate, chondroitin 6-sulhate |
| | MPS IV B (Morquio B) | β-Galactosidase | Keratan sulphate |
| | MPS VI | 4Acetyl galactosamine - sulphatase (arylsulphatase B) | Dermatan sulphate |
| | MPS VII (Sly) | β-Glucuronidase | Chondroiotin 6-sulphate, dermatan sulphate, heparan sulphate, |
| | MPS IX (Natowicz) | Hyaluronidase | Hyluronan |
| **Sphingolipidoses** | Fabry | α-Galactosidase A | Globotriasylceramide |
| | Farber | Acid ceramidase | Ceramide |
| | Gangliosidosis GM1 (Types I, II, III) | GM1-β-galactosidase | GM1 ganglioside,Keratan sulphate, oligos, glycolipids |
| | Gangliosidosis GM2, Tay-Sachs | β-Hexosaminidase A | Glycolipids, oligos, GM2 ganglioside |
| | Gangliosidosis GM2, Sandhoff | β-Hexosaminidase A + B | Oligos, GM2 ganglioside |
| | Gaucher (Type I, II, III) | Glucosylceramidase | Glucosylceramide |
| | Krabbe | β-Galactosylceramidase | Galactosylceramide |
| | Metachromatic leucodystrophy | Arylsulphatase A | Sulphatides |
| | Niemann-Pick (type A, type B) | Sphingomyelinase | Sphingomyelin |
| **Glycoproteinoses** | Aspartylglicosaminuria | Glycosylasparaginase | Aspartylglucosamine |
| | Fucosidosis | α-Fucosidase | Glycoproteins, glycolipids, Fucoside-rich oligos |
| | α-Mannosidosis | α-Mannosidase | Mannose-rich oligos |
| | β-Mannosidosis | β-Mannosidase | Man(β1 → 4)GlnNAc |
| | Schindler | N-acetylgalactosaminidase | Sialylated/asialoglycopeptides, glycolipids |
| | Sialidosis | Neuraminidase | Glycopeptides, oligos |
| | Glycogenosis II/Pompe | α1,4-glucosidase (acid maltase) | Glycogen |
| | Wolman/CESD | Acid lipase | Cholesterol esters |

| | | | |
|---|---|---|---|
| **Transmembrane protein defect transporters** | Sialic acid storage disease; infantile form (ISSD) and adult form (Salla) | Sialin | Sialic acid |
| | Cystinosis | Cystinosin | Cystine |
| | Niemann-Pick Type C1 | Niemann-Pick type 1 (NPC1) | Sphingolipids and cholesterol |
| | Niemann-Pick, Type C2 | Niemann-Pick type 2 (NPC2) | Sphingolipids and cholesterol |
| **Lysosomal enzyme protection defect** | Galactosialidosis | Protective protein cathepsin A (PPCA) | Sialyloligosaccharides |
| **Non-enzymatic lysosomal protein defect** | Gangliosidosis GM2, activator defect | GM2 activator protein | GM2 ganglioside, oligos |
| | Metachromatic leucodystrophy | Saposin B | Sulphatides |
| | Krabbe | Saposin A | Galactosylceramide |
| | Gaucher | Saposin C | Glucosylceramide |
| **Trafficking defect in lysosomal enzymes** | Mucolipidosis IIα/β, IIIα/β | GlcNAc-1-P transferase | GAGs, Oligos lipids |
| | Mucolipidosis IIIγ | GlcNAc-1-P transferase | Oligos, GAGs, lipids |
| **Structural Proteins** | Danon | Lysosome-associated membrane protein 2 | Glycogen and cytoplasmatic debris and |
| | Mucolipidosis IV | Mucolipin | Lipids |
| **Polypeptide degradation defect** | Pycnodysostosis | Cathepsin K | Bone proteins |
| **Neuronal ceroid lipofuscinoses (NCLs)** | NCL 1 | Palmitoyl protein thioesterase (PPT1) | Saposins A and D |
| | NCL 2 | Tripeptidyl peptidase 1 (TPP1) | Subunit c of ATP synthase |
| | NCL 3 | CLN3, lysosomal transmembrane protein | Subunit c of ATP synthase |
| | NCL 5 | CLN5, soluble lysosomal protein | Subunit c of ATP synthase |
| | NCL 6 | CLN6, transmembrane protein of ER | Subunit c of ATP synthase |
| | NCL 7 | CLC7, lysosomal chloride channel | Subunit c of ATP synthase |
| | NCL 8 | CLN8, transmembrane protein of endoplasmic reticulum | Subunit c of ATP synthase |
| | NCL 10 | Cathepsin D | Saposins A and D |
| **Post-translational processing defect** | Multiple sulphatase deficiency | Multiple sulphatase | GAGs, Sulphatides, glycolipids |

## 1.1.2. Molecular basis of LSDs

Due to a vast nature of undigested molecules (mucopolysaccharides, glycoproteins, lipids or glycogen) across all LSDs and the secondary pathological events following the abnormal substrate accumulation, the understanding of the molecular basis underlying this family of diseases needs to be undertaken one syndrome/pathology at a time (Platt

et al., 2018; Winchester, 2005; Winchester et al., 2000). However, the family of LSDs shares one common denominator, which is the 'lysosome' (Hers, 1963; Vellodi, 2005).

Lysosomes due to their pivotal role in the breakdown and recycling of macromolecules serve as a metabolic centre that controls amino acids and ion homeostasis, calcium signalling and nutrient sensing (Medina & Ballabio, 2015; Settembre et al., 2013; Todkar et al., 2017). They fuse with phagosomes and/or autophagosomes and the plasma membrane leading to release of molecules into extracellular space. Some of these molecules are by-products of intracellular metabolism, others come from the breaking down of bacterial/pathogens proteins being then shown by MHC II complex to immune cells, and others being signalling molecules i.e. hormones or other activators (Nakagawa & Rudensky, 1999; Settembre et al., 2013). Lysosomes, despite for a long time being considered as a cellular damper, in light of new research play a crucial role in cell-cell interaction taking part in immune responses to infection and maintaining general homeostasis of cells (Medina & Ballabio, 2015; Settembre et al., 2013).

The lysosomal storage of non-degraded substrates sets off not only a lysosome enlargement and mechanical stress, but also triggers several other secondary intracellular mechanisms/events, which lead to cell damage and cell death (Ballabio & Gieselmann, 2009). Enlargement of autophagosomes, release of lysosomal contents of low pH containing acid hydrolases into cellular matrix damaging cellular homeostasis, inflammation endoplasmic reticulum stress, distorted lipid trafficking are only some of the examples of indirect damage caused by mutations in genes responsible for lysosomal function (Ballabio & Gieselmann, 2009; Medina & Ballabio, 2015; Settembre et al., 2013; Vitner et al., 2010; Vitner et al., 2014).

## 1.2. Pompe disease (PD)

Pompe disease (PD), also known as acid maltase deficiency or glycogen storage disease type II (GSDII) is caused by mutations in the gene encoding the lysosomal enzyme acid alpha-glucosidase (GAA), which is responsible for the cleavage of glycogen into glucose and more specifically for the hydrolysis of the $\alpha$-1,4- and $\alpha$-1,6-glycosic bonds in

glycogen (Angelini & Engel, 1973; Beratis et al., 1978; Hers, 1963; Nicolino et al., 1997). The hallmark of PD is a permanent absence or severe reduction of GAA enzyme activity that leads to accumulation of glycogen in lysosomes in various tissues (Lim et al., 2014a; Platt et al., 2018), with primarily clinical symptoms mainly manifested in skeletal and smooth muscle, the heart and CNS (Boentert et al., 2016; van der Beek et al., 2012) . GSDII was first described by a Dutch pathologist, Johannes Cassianus Pompe in 1932 who studied a post-mortem case of 7 month old infant with a cause of death indicated as idiopathic hypertrophy of the heart and general muscle weakness (Boentert et al., 2016; Hers, 1963). The main observation of Dr Pompe was that the infant's symptoms were linked to immensely enlarged vacuoles containing stored glycogen across all types of tissues in the body (Boentert et al., 2016; Lim et al., 2014a). This observation was of paramount importance and lies at the foundation of lysosomal storage disorders with many more syndromes yet to be discovered with different pathologies and various gene mutations at their bases (Boentert et al., 2016).

**1.2.1 GAA *gene and* the distribution of disease-associated mutations in GAA protein domains.**

The GAA gene has been localized to chromosome 17q21-23 and is approximately 28 kb long, comprising of 20 exons.  The first ATG is located in exon 2, as the entire exon 1 in untranslated. The GAA cDNA is 3.6 kb long giving rise to a precursor peptide of 952 amino acids with molecular weight of 102 kDa (Roig-Zamboni et al., 2017).

To date there are 582 known alterations in the *GAA* gene (Peruzzo et al., 2019), causing either complete absence of GAA enzyme or a very limited residual activity  (Pittis et al., 2008; van Gelder et al., 2015). The highest proportion of mutations described in *GAA* gene are missense mutations (51%), followed by small deletions/frame shifts (14.9%). Third place are represented by splicing variants (12.7%), then nonsense mutations with 8% of all mutations, followed by 6% representing small insertions/duplications. Major insertions/deletions and small indels are represented by 3.3% and 2.2% respectively.

Aforementioned mutations are mainly confined to exons, however intronic mutations can also lead to progression of Pompe disease over time by affecting splicing efficiency of normal GAA transcript thereby reducing fully functional copies of GAA protein. These variable levels of residual GAA activity can lead to delayed phenotypic display of Pompe disease.

The mutational range across *GAA* gene is extremely heterogeneous. Figure 1.1 A depicts diverse distributions of only some novel variants reported in the Pompe Registry ( https://www.pompevariantdatabase.nl/pompe_mutations_list.php?orderby=aMut_ID 1 ), followed by Figure 1.1 B illustrating subsequent affected domains in the GAA protein (GAA protein domains are described in 1.2.2 section).
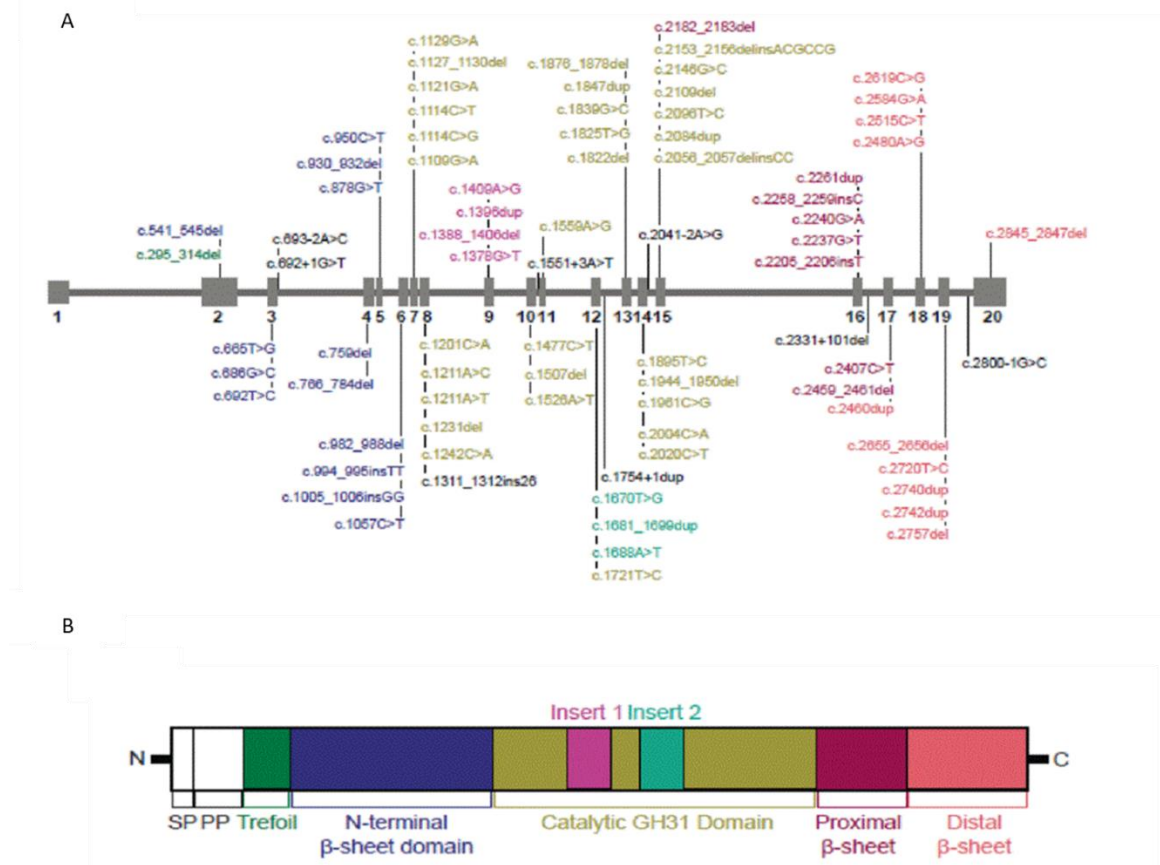


Figure adapted from *Reuser et al.*

*Figure 1.1 Sites of novel mutations in GAA gene. A) Spectrum of novel mutations reported in the Pompe Registry. B) Domains of the GAA protein with corresponding color-coded mutations*

Nevertheless, a precise correlation between a particular mutation and the clinical presentation and progression of the disease is difficult to establish in PD as for most genetic disorders. In general, majority of Cross Reactive Immunologic Material (CRIM) negative patients have mutations in form of frameshift and substitution (nonsense) variants leading to premature stop codons and synthesis of non-functional GAA proteins and lack of any residual enzyme activity and rapid onset of the disease. In contrary CRIM positive patients show substitutions (missense) and deletions mutations leading to production of active, to some degree, form of GAA leading to delayed, various in time, displayed onset of PD. Figure 1.2 illustrates disease-associated mutations linked to functional GAA protein domains, and illustrates that the most severe mutations are located, most commonly, to the catalytic domain of the enzyme and preceding it in the N-terminal β-sheet domain.
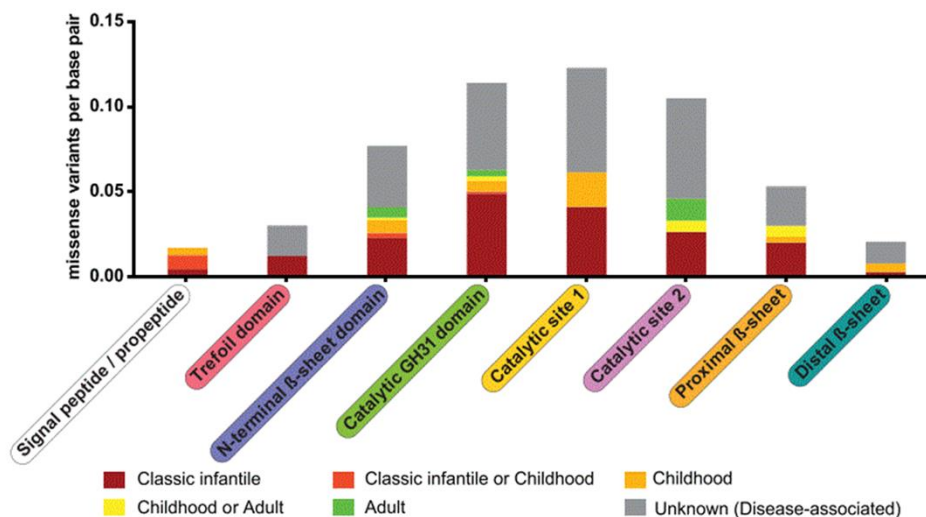


Figure adapted from *Niño et al.*

*Figure 1.2. The location of disease-related mutations in GAA protein domains according to prevalence and clinical severity score. Highest frequency variance in GAA is observed*

*in catalytic domains and N-terminus β-sheet domain. Figure adapted from Niño et al.(Nino et al., 2019).*

Furthermore, Kroos *et al.* proposed a classification of GAA mutations based on their severity and organized them in six classes identified by letters, where A represents "very severe", B used for "potential less severe", C "less severe", D represents "potentially mild", E "presumably non-pathogenic" and at last F for "non-pathogenic" (Kroos et al., 2008). Class A encompasses the most severe forms of PD diseases with neither GAA enzyme protein detected nor residual activity measured; this class of mutations is CRIM-negative (Kroos et al., 2008). This group is the most likely to develop IgG antibody against any treatment involving exogenous form of recombinant human GAA (rhGAA) (Desai et al., 2019), because the immune system has not yet been tolerised to GAA antigen and therefore reacts to any exogenous GAA expression. Classes B, C and D fall into CRIM-positive category as GAA protein and its activity are detectable and can range from 1 to 30% of the wild type levels (Kroos et al., 2008).These groups due to residual endogenous GAA protein are much less likely to develop any IgG immune response to exogenous rhGAA and have much better clinical prognosis than the CRIM-negative group (Desai et al., 2019). Class E represents a level of GAA protein and its activity between 30-60% of wild type, while values higher than 60% are found in group F (Kroos et al., 2012; Kroos et al., 2008). To date, class A and B encompass 183 and 139 different mutations respectively, while 30 mutations are described in class C, 22 mutations in class D and 6 mutations in class E. In addition, class F spans over 93 different sequence variations and other 85 mutations have been found but their severity levels remain unclear (Kroos et al., 2012).

### 1.2.2. Alpha acid glucosidase (GAA)

*GAA* gene is 28kb long and covers 20 exons, with first exon untranslated followed by first ATG in exon 2 (Roig-Zamboni et al., 2017). Three common allelic variants of *GAA* have been described coding for three alloenzymes: the most common GAA*1, GAA*2 and GAA*4 (Dasouki et al., 2014). The *GAA* cDNA is 3.6kb in length (Martiniuk et al., 1986) and gives rise to 952 amino acid precursor peptide with a molecular weight of

105kDa translated in the endoplasmic reticulum (ER) (Dasouki et al., 2014). The juvenile form of GAA protein consists of several different domains, depicted on Figure 1.3, from which catalytic GH31 domain containing 2 catalytic sits (described as catalytic site 1 and catalytic site 2) playing a critical role in recognition and interaction with its substrate. Furthermore, the trefoil Type-P domain has been identified as potential additional substrate-binding pocket (Reuser et al., 2019). Furthermore, several β-sheet domains are located across the protein providing confirmational support and active binding to the substrate (Reuser et al., 2019).
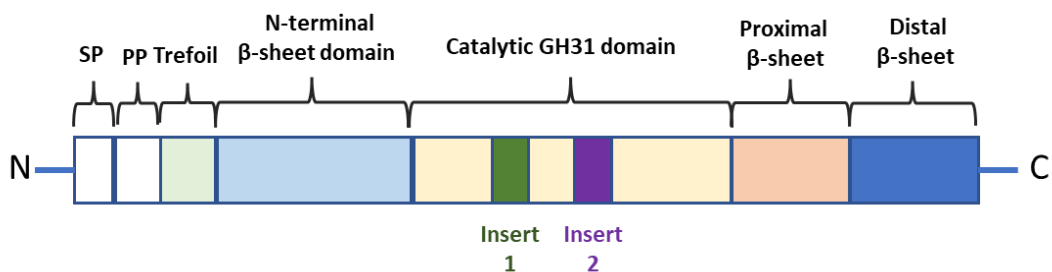


*Figure 1.3. **Schematic representation of GAA protein.** Juvenile form of GAA protein starting from N-terminus in consecutive domains: SP-signal peptise, PP-propeptide, Trefoil Type-P domain, N-terminus β-sheet domain, Catalytic GH31 domain containing two catalytic sides (insert 1 and 2), Proxima β-sheet domain and Distal β-sheet domain- End C-terminus.*

The ER is  the first place where GAA protein is modified by addition of high-mannose oligosaccharides in seven different locations (Hermans et al., 1993), in a process known as glycosylation, increasing the molecular weight of GAA to 110kDa (van der Horst et al., 1987). The enzyme in question is then transferred to Golgi complex where mannose residues are phosphorylated (Hasilik & Neufeld, 1980) and consequently loaded in lysosome-directed vesicles using mannose 6-phosphate receptor (M6PR). Once, GAA is delivered to late endosomes/lysosomes the M6PR-ligand complex separates due to a low pH and M6PR is recycled back to Golgi complex to serve for another round of lysosomal protein transport (Braulke & Bonifacino, 2009). In late endosome/lysosomes

GAA precursor undergoes a sequence of proteolytic cleavages resulting in a 95kDa intermediate, which is processed at the C- terminus to 76kDa protein followed by a proteolytic cleavage at N-terminus that finally releases the most mature GAA form of 70kDa in size and with full catabolic activity (Hasilik & Neufeld, 1980). This proteolytic cascade is known as maturation process and amplifies GAA specificity and activity toward glycogen about 10-fold compared to immature GAA protein (Hasilik & Neufeld, 1980). Biosynthesis, maturation and lysosomal transport is presented on Figure 1.4.
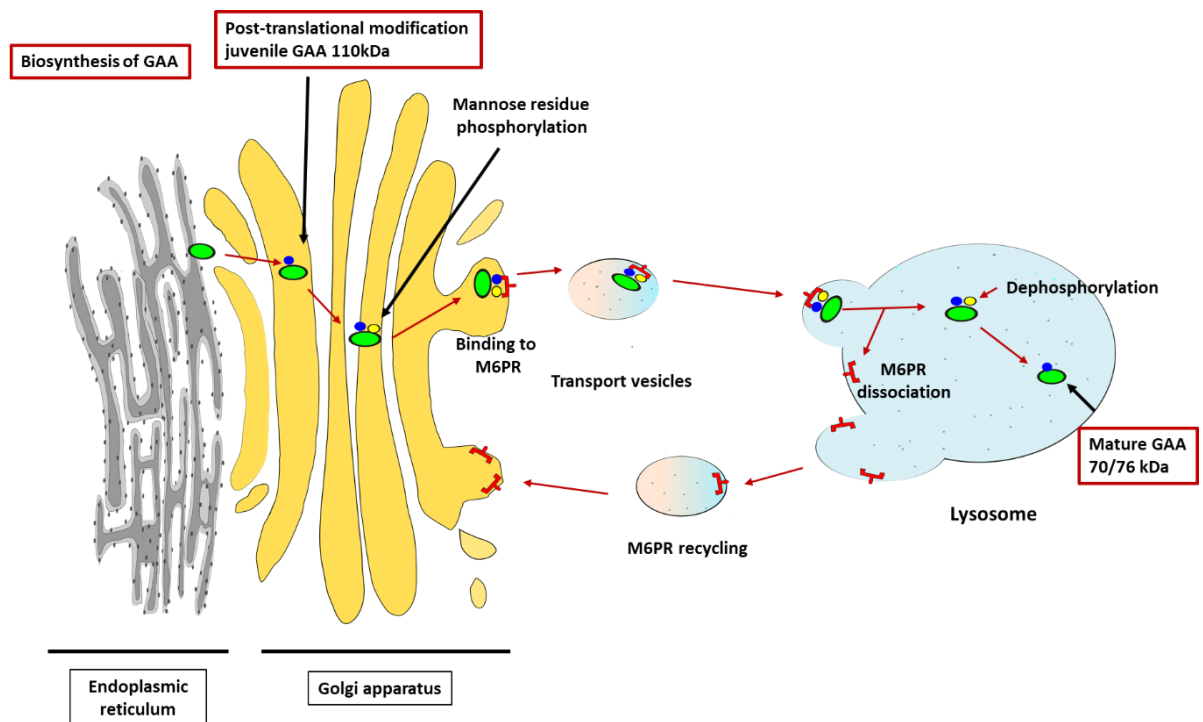


*Figure 1.4. Biosynthesis, maturation, and lysosomal transport of acid alpha glucosidase. Juvenile form of GAA 110 kDa is synthesised at the endoplasmic reticulum where proteins are being correctly folded and modify with functional groups i.e., oligosaccharides, methylation ect. Next, GAA is transported to Golgi where undergoes phosphorylation to its mannose groups (M6P). M6P groups are being recognised by M6PR follow by entrance of GAA in transport vesicle. Docking to late endosomes allow GAA to maturate in acidic pH to 70/76 kDa mature hydrolases and in turn M6PR is being recycled back to Golgi where can serve to transport other lysosomal hydrolases.*

### 1.2.3 Glycogen

Glycogen embodies the foremost polysaccharides storage found primarily in animals, bacteria and fungi (Adeva-Andany et al., 2016). In animals glycogen is mainly present in liver (as a main storage of carbohydrates molecules) and muscles, with a substantial storage also found in brain (Caballero et al., 2003). The synthesis of glycogen (glycogenesis) is crucial for the retention of surplus of glucose and, on contrary, its degradation (glycogenolysis) provides energy/fuel during time of need i.e., physical activity, exercises or starvation (Adeva-Andany et al., 2016; Caballero et al., 2003). The importance of glycogen is not related only to an energy source, but it is also very important as a sensor molecule that aid metabolic adaptations associated with intracellular osmolality, training adaptation and improving endurance performance (Burke et al., 2017; Philp et al., 2012).

Glycogen molecule is a branched carbohydrate formed of up to 60,000 of α-D-glucose molecules connected in linear manner by α-1,4 glycosidic bonds while the branched points are created by α-1,6 glycosidic linkages, Figure 1.5 (Chikwana et al., 2013). The branching of glycogen molecule increases its solubility and reduces the osmotic effect of the particle in question (Caballero et al., 2003). A size of glycogen molecule varies through tissues ranging from 10-44nm in circumference in muscles to approximately 110-290nm found in liver (Gibbons et al., 2002).
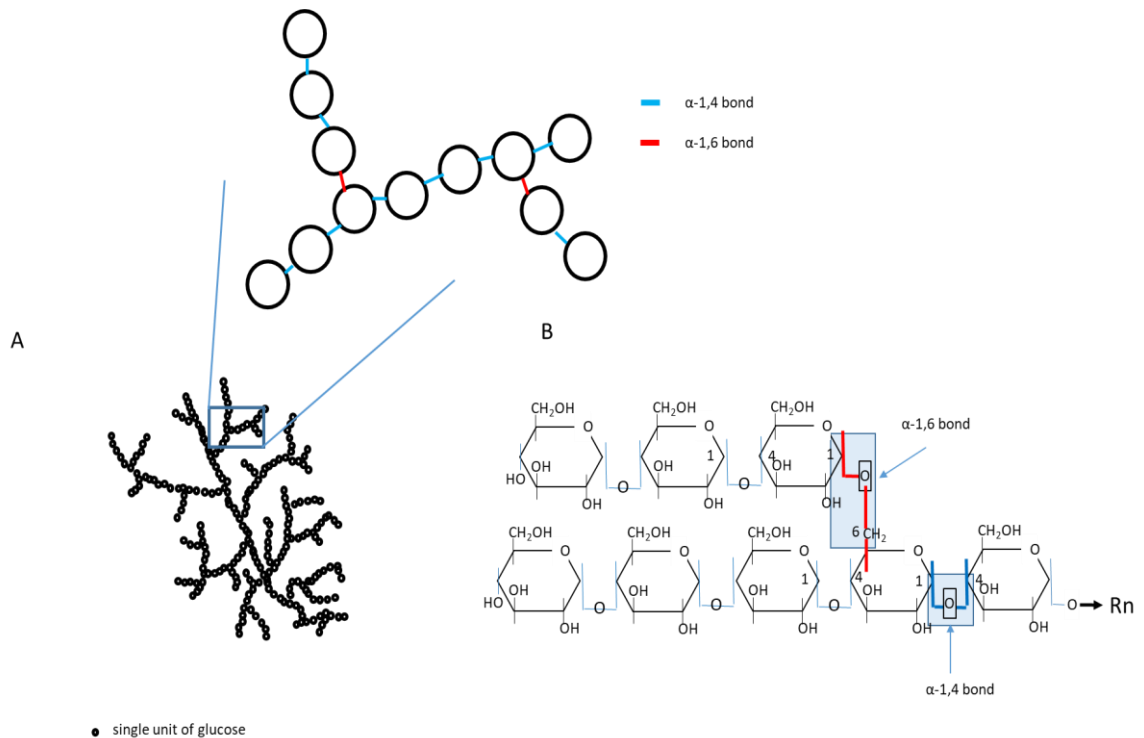
*Figure 1.5 Structure of Glycogen.* A) Biochemical composition of glycogen with *α-1,4 linkage* between linear units of glucose and *α-1,6 bonds* allowing for a formation of branches. B) Branched glycogen made up of many α-D-glucose units.

Both glycogenesis and glycogenolysis conducted in liver are tightly regulated processes through a hormonal control, mainly of insulin and glucagon/adrenalin, respectively. Glycogenolysis in muscle, on top of the hormonal regulation, is also influenced by local stimuli requiring/inducing an instant local energy demands/supply i.e., extensive physical activity/exercises, stress (Ellingwood & Cheng, 2018).

Glycogenesis is a complexed multi-enzymatic process requiring several subsequent enzymes to convert initial glucose molecules into a glycogen particle. On the other hand lies the more straightforward process of glycogenolysis, which occurs in two distinctive ways with one taking place in cytosol and the other in the lysosomes (Johnson, 1992). The cytosolic glycogenolysis is undertaken by two enzymes: glycogen phosphorylase (GP), which cleaves α-1,4 glycosidic bonds releasing glucose from the linear glycogen chain, and glycogen debranching enzyme (GDE), which breaks α-1,6 glycosidic linkages at the branching points. Lysosomal glycogenolysis is carried out by the acid α-glucosidase (GAA) alone (Lim et al., 2014b) (Figure 1.6).
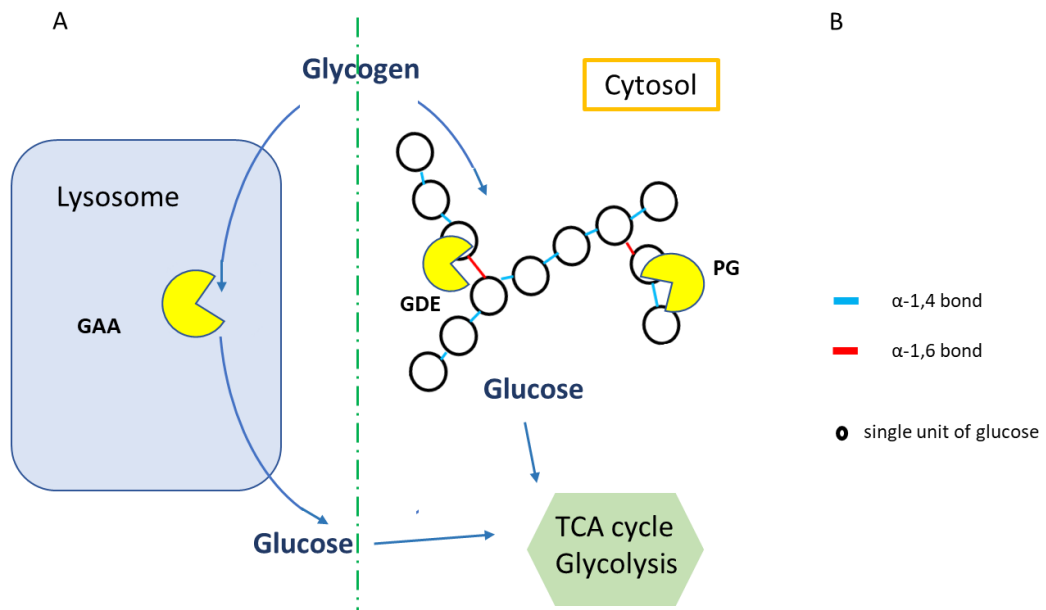
**Figure 1.6. Glycogen degradation pathway**. **A)** *Lysosomal glycogen broken down to single molecules of glucose by hydrolysis of the α-1,4- and α-1,6-glycosic bonds by acid alpha glucosidase (GAA).* **B)** *Cytoplasmic degradation of glycogen to glucose by two enzymes; glycogen debranching enzyme (GDE), breaking α-1,6 glycosidic linkages and glycogen phosphorylase (GP), cleaving α-1,4 glycosidic bonds. Molecules of glucose enter Tricarboxylic Acid Cycle (TCA) and/or Glycolysis.*

Glucose released from glycogen stored in liver and secreted into the blood stream maintains a constant systemic level of the fuel molecule, while muscle released glucose, either in cytosol or/and lysosomes, remain sourced locally (Adeva-Andany et al., 2016). The presence of two distinctive pathways processing glycogen within one cell is not fully understood and more intriguingly dysfunctional enzyme (due to a genetic mutation or/and acquired errors) from one path cannot be replaced by functional enzymes from the other one (Ball et al., 2011). This in consequence induces a formation of glycogen storage burden, a prototype of glycogen storage disorder (GSD), resulting in severe cellular dysfunctions followed by organ malfunction and inevitable significant clinical consequences that may shorten the life span of an affected individual (Ellingwood & Cheng, 2018).

## 1.2.4. Clinical manifestations of PD

Clinical manifestations of various mutations in *GAA* gene are very heterogeneic and vary from a very early onset in an infant life - Infant Onset Pompe Disease (IODP) - through childhood, juvenile or adulthood -Late Onset Pompe Disease (LOPD) (Ebbink et al., 2018; van der Ploeg & Reuser, 2008). IOPD phenotype is characterised by either the absence of GAA protein or the lack of residual GAA activity (Ebbink et al., 2016), where clinical manifestation is significantly muscular related, with proximal muscle weakness, macroglossia, hepatomegaly, severe hypertrophic cardiomyopathy, and respiratory function being critically affected (van der Ploeg & Reuser, 2008). Infants born with IOPD, if not treated in time, suffer from an inevitable death due to cardiac and respiratory insufficiency within the first 12 months of age (van den Hout et al., 2003). IOPDs are also sub-classified as classic or non-classic PD according to the presence or absence of cardiomyopathy in patients, respectively (Case et al., 2012; Cupler et al., 2012). Patients with LOPD show various levels of residual GAA activity leading to a slower PD progression. However, patients show deterioration of their respiratory system and CNS that can cause cerebral aneurysm and rigid spine syndrome. In addition, LOPD patients show a progressive limb-girdle myopathy, genioglossus, bilateral ptosis (El-Gharbawy et al., 2011), scoliosis (Kobayashi et al., 2010) and the respiratory deficiency is described as the main cause of fatality in this group (Hagemans et al., 2005). In LOPD patients severe cardiomyopathy is not present (El-Gharbawy et al., 2011), however cardiac involvement includes dilation of the ascending aorta, left ventricular hypertrophy and also Wolff-Parkinson-White condition that leads to persistent tachycardias (Preisler et al., 2013).

## 1.2.5. Diagnosis of PD

The majority of LSDs including PD are relatively complicated to diagnose as they do not present clear-cut clinical manifestations in a first glance and can be easily mistaken for other medical conditions, especially in newborn patients (Manganelli & Ruggiero, 2013). The onset of PD and the pathological clinical symptoms are critically dependent on the level of residual enzymatic activity (Gritti, 2011).

To date there are two ways to diagnose PD with one of them being a clinical evaluation followed by laboratory testing (Disease et al., 2006). Due to the infrequency of the condition and relatively generic nature of the phenotypic features, that only in cumulative symptoms lead to suspicion of the disease in question, the identification of PD poses a diagnostic dilemma (Disease et al., 2006).

Alongside a clinical evaluation, a laboratory diagnostic test can be run to evaluate the level of GAA enzyme in patients' blood sample or from muscle or skin biopsies (Disease et al., 2006). Currently, GAA enzymatic activity assay is processed with addition of acarbose to inhibit maltase-glucoamylase (MGA), an enzyme that was found to interfere with the outcomes of the assay (Li et al., 2004). The employment of this specific inhibitor of MGA improved accuracy of GAA activity assays and thereby of the diagnosis of PD (Disease et al., 2006).

In addition, newborn screening programmes (NBSPs) to a number of serious conditions, including Pompe disease, were developed in several countries i.e., Taiwan, Japan and United States of America (Burton et al., 2020; Hwu & Chien, 2020). NBSPs aim to reduce delay in diagnosis of life-threatening disorders and are normally performed withing first 48 hrs after birth (Hwu & Chien, 2020; Tang et al., 2020). PD screening is primarily based on GAA enzymatic activity from a new-born's sample of blood and if positive a second testing is performed to confirm or reject primary testing results. Secondary assessment is typically a genetic test performed to determine possible mutation within *GAA* gene (Burton et al., 2020; Saich et al., 2020; Smith et al., 2020).

### 1.2.6. Cellular pathogenesis of PD

The pathological accumulation of the glycogen, due to the inability of GAA to break it down has been considered the main culprit of the pathology seen in PD. (Platt et al., 2018).

Although the clinical manifestations of PD are well characterised, the pathogenesis of this disease is far more complex than it was initially described (DiMauro & Spiegel, 2011). It is well documented that the initial trigger of Pompe disease is the build-up of

glycogen inside lysosomes in all types of tissues, however in the light of new studies the role of autophagy becomes a focal point in the cellular pathology of Pompe disease (Fukuda et al., 2006; Malicdan & Nishino, 2012; Raben et al., 2012). Bulky accumulations of autophagic debris mainly in type 2 skeletal muscle fibres, were present in patients with PD and also in *Gaa -/-* mouse models (Fukuda et al., 2006; Raben et al., 2012). Furthermore, studies revealed that in Pompe patients a concentration of autophagosomes in skin fibroblast and skeletal muscle was associated to an abnormal CI-M6PR plasma membrane distribution (Fukuda et al., 2006) and cellular trafficking, which then was linked to reduced response to ERT treatment (Raben et al., 2008). Overall, the deficiency of GAA enzyme in patients with Pompe is not only limited to the lysosomes but this pathology reaches beyond muscles' glycogen accumulation and extend to a range of vesicular structures linked to lysosomes including recycling endosomes, early endosomes and trans-Golgi complex (Orth & Mundegar, 2003). Furthermore, individuals affected by LOPD show increased total of Insulin-like growth factor 1 (IGF-1) and Insulin-like growth factor binding protein-3 (IGFBP -3), which indicate cellular difficulty in nutrient sensor interactions based on energy deficiency in skeletal and smooth muscle (Pascual & Roe, 2013). Nevertheless, the pathogenesis of PD is still poorly understood and requires further investigations to elucidate the cellular/systemic pathology caused by the accumulation of glycogen in lysosomes (Dasouki et al., 2014).

## 1.3. Treatments for LSDs with particular focus on Pompe disease

A number of LSDs are currently treatable: for instance, the volume of build-up material inside lysosomes can be diminished by substrate reduction therapy (SRT) (Coutinho et al., 2016); active enzyme can be taken up through exogenous ERT (Kishnani & Beckemeyer, 2014) or pharmacological compound can improve/stabilise the faulty lysosomal enzyme such as pharmacological chaperone in Pharmacological Chaperone Therapy (PCT )(Beck, 2010a; Smid et al., 2010). In addition, gene therapy utilising LV in hematopoietic stem cells (HSCs) (Stok et al., 2020; van Til et al., 2010) or AAV (Salabarria et al., 2020; Samulski & Muzyczka, 2014) might be also potential treatment to alleviate

LSDs symptoms. However, due to the great heterogeneity of LSDs their treatment results range from extremely successful to negligible effect (Platt et al., 2018).

### 1.3.1. Substrate reduction therapy (SRT)

SRT approach aims to reduce or slow down the assembly of lysosomal storage materials rather than cause their degradation (Coutinho et al., 2016). For instance, Gaucher disease can be slowed down and its symptoms be alleviated by the application of the iminosugar N-butyldeoxynojirimycin which strongly inhibits ceramide glucosyltransferase (GlcCer synthase), the enzyme responsible for the biosynthesis of the storage material glucosylceramide, which build-up causes Gaucher disease type I (Coutinho et al., 2016; Cox et al., 2000; Hollak et al., 2009; Poole, 2014). This substrate reducing molecule is commercially available under the name Miglustat and it is orally administrated (Platt & Lachmann, 2009). In addition to treat Gaucher disease, Miglustat can also improve symptoms in Niemann-Pick disease type C (NPC), a neurovisceral disease (Patterson et al., 2007; Pineda et al., 2018) where NPC1 or NPC2 genes, encoding proteins responsible for cellular lipid trafficking, are mutated leading to the accumulation of sphingosine, glycosphingolipids, sphingomyelin, and cholesterol predominantly in CNS (Beck, 2010b). Due to its small molecular weight Miglustat can cross the blood-brain barrier (BBB) and, as shown in NPC registry study, patients receiving Miglustat for over 2 years stabilise or even enhance their cognitive functions (Barton et al., 1991) as well as improve swallowing, which in turn reduces incidents of death caused by aspiration pneumonia (Lyseng-Williamson, 2014; Walterfang et al., 2012). A number of potential SRT drugs are currently being tested in phase II clinical trials including Ibiglustat, allosteric inhibitor of glucosylceramide synthase that lowers glycosphingolipid synthesis, for Gaucher and Fabry disease and also genistein, soy-derived isoflavone reducing proteoglycan biosynthesis, for MPS III (Piotrowska et al., 2006).

Based on the success in other lysosomal diseases (Kishnani, Hwu, et al., 2006), SRT has also been tested in animal models of PD (Lim et al., 2014a). There are two major players in the biosynthesis of glycogen in cells: glycogen synthetase (GYS) and glycogenin (GYG)

(Preiss, 2019). The inhibition of these two enzymes by muscular injection of AAV1 expressing shRNAs showed significant reduction in cellular and lysosomal glycogen build-up and also reduction of lysosomal dimension in *Gaa -/-* mouse model (Douillard-Guilloux et al., 2008). In addition, the reduction of expression of GYS1 (muscle isoform) in *Gaa -/-* animals led to cardiac muscle correction and lessened glycogen accumulation in muscle tissues (Douillard-Guilloux et al., 2008). In contrast, the null mutation of GYS1 was associated to cardiac failure in glycogen storage disease type 0 (Kollberg et al., 2007), therefore not a systemic, but a muscle specific reduction of GYS1 could be the safest option to attempt to alleviate PD clinical symptoms (Kollberg et al., 2007).

### 1.3.2. Stop-codon read-through

In a number of LSDs a nonsense mutation, a premature stop codon, leads to an early termination of protein translation and to the formation of a truncated and not functional enzyme (Brooks et al., 2006). It is known that gentamicin, a low molecular mass drug, can force a read-through of premature stop codon leading to the formation of an enzyme protein with subnormal or normal enzymatic activity (Brooks et al., 2006). This nonsense mutation suppression therapy is being successfully tested in cystic fibrosis and Duchenne muscular dystrophy (Keeling & Bedwell, 2011). In addition, *in vitro* studies of MPS I caused by nonsense mutations have also shown the benefit of gentamicin treatment in restoring enzymatic activity (Clarke et al., 2009). However, the use of gentamicin in long-term treatment might be restricted due to its known toxicities including kidney damage and hearing loss (Clarke et al., 2009). Research for new, safe stop-codon read through drugs is desired.

### 1.3.3. Exon-skipping

The sequence of genes coding for proteins are made of introns and exons, with both of these structures being consecutively transcribed into pre-mRNA (Papasaikas & Valcarcel, 2016). However, only exons carry a vital information for production of functional proteins and introns are nonsense information in genetic code that will be removed in the maturation of mRNA in the process called constitutive splicing carried

by the spliceosome (Galej et al., 2014)(Figure 1.7 A). In human genome single pre-mRNA can code for many different proteins depending on a specific cell type, environmental stress etc. through the process called alternative splicing where some exons are being removed and others kept generating several protein isoforms (Figure 1.7 B).

Pre-mRNA splicing process can be affected by mutations in genes falling in LSD scope leading to incorrect messenger RNA being created. Across LSDs mutations in splice acceptor consensus sequence or splice donor range between 5-19 % and they are LSD-dependent (Abramowicz & Gos, 2019; Anna & Monika, 2018; Aung-Htut et al., 2020; Baralle & Buratti, 2017; Jaganathan et al., 2019). A single mutation in the coding gene can alter the splicing process resulting with incorrect parts of exons combined together and shifting of reading frame which thereby cause a formation of a dysfunctional protein (Dardis & Buratti, 2018). Mutations causing splicing errors include c.894G>A in lysosomal acid lipase (LIPA) gene, c.459+5A>G in β-hexosaminidase A (HEXA) gene, α-galactosidase A (GLA) gene 639+919G>A, and related to Pompe, c.-32-13T>G in *GAA* gene (Dardis & Buratti, 2018).

Currently, small molecules modulating expression of some splicing proteins, modified small nuclear RNAs (U1snRNA) or antisense oligonucleotides (AONs), have been tested in vitro to mask the mutation site and allow correct splicing (Dardis & Buratti, 2018). Most promising and already approved by the American Food and Drug Administration (FDA) is AONs therapy for the treatment of Duchenne muscular dystrophy (DMD) (Finkel et al., 2016). While in *in vitro* study of LSDs, the U1snRNA was tested in the attempt to rescue GLA gene mutation resulting in partial enzymatic activity restoration (Ferri et al., 2016). In addition, a very promising discovery has been made by studying the c.-32-13T>G GAA mutation in patient's derived myocytes using AONs (Goina et al., 2017) which substantially rescued the normal splicing of GAA mRNA leading to a significant reduction in glycogen in patient myoblast (Ferri et al., 2016).
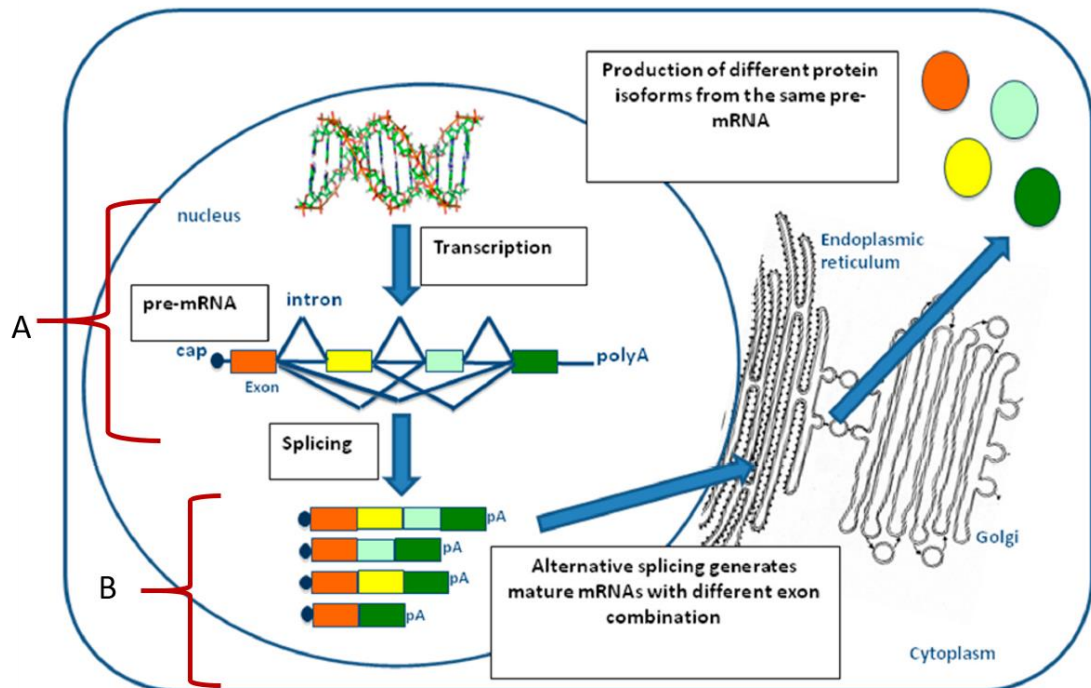
Figure adapted from *Daris and Buratti*.

***Figure 1.7. Splicing process.*** *A Transcription of a gene into pre-mRNA molecule, introns and exons alternatively placed alongside. B. Splicing process of the pre-mRNA with subsequent removal of introns. Alternative splicing shown by matching different exons (colour blocks) with each other from a single pre-mRNA causes formation of different mRNA leading to the assembly in ER of different proteins. Figure adapted from Daris and Buratti (Dardis & Buratti, 2018).*

### 1.3.4. Pharmacological chaperone therapy (PCT)

All living cells have mechanisms to control the exact folding of newly synthesised peptides/enzymes and any misfolded proteins are being broken down and recycled (Goemans et al., 2014). Unique proteins so-called chaperones (calnexin, heat shock proteins) can facilitate the correct folding of proteins (Saibil, 2013). In the case of LSDs, many genes encoding for lysosomal enzymes carry mutations resulting in unstable/misfolded proteins, which in turn lead to a lack of catalytical activity or a short protein half-life (Parenti, Andria, & Valenzano, 2015).

PCT can extend the half-life of mutant enzymes and recover partially their residual enzymatic activity (Beck, 2010b; Haneef & Doss, 2016; Smid et al., 2010). Currently, PCT is used in clinical trials for treatment of Fabry, Gaucher and Pompe disease (Gritti, 2011). However, the only approved and available chaperone medicine, up to date, is Migalastat for the treatment of Fabry's type I and II disease (Markham, 2016; Parenti, Andria, & Valenzano, 2015). Migalastat is a pharmacological chaperone, which in Fabry's disease stabilises a mutant form of α-galactosidase A enzyme and facilitates its trafficking from ER to lysosomes (McCafferty & Scott, 2019) improving clinical features seen in Fabry's patients such as ventricular hypertrophy and decreases the severity of indigestion, diarrhoea and reflux (Germain et al., 2016; Schiffmann et al., 2018). This chaperone molecule has also an ability to cross blood-brain barrier and, therefore, has the potential to treat neuropathic manifestations in LSDs (Platt et al., 2018).

A study using Pompe patients' fibroblasts and the combined therapy of chaperone N-butyldeoxynojirimycin (NB-DNJ) and ERT with rhGAA showed much better outcomes in rescuing the GAA activity and reducing glycogen level than the sole administration of each therapy (Porto et al., 2009). This chaperone is not approved yet for clinical trials and requires further investigation. However, the evidence suggest that combined therapy might bring additional benefits to stabilise rhGAA protein by extending its half-life and allowing greater bioavailability (Porto et al., 2009).


### 1.3.5. Enzyme Replacement Therapy

In many LSDs, gene mutations lead to the production of lysosomal enzymes lacking their catabolic activity or with very limited capability to break down their substrate (Vellodi, 2005). ERT utilises the ability of cells to take up macromolecules, such as enzymes, by endocytosis and traffic those into lysosomes, where functional exogenous enzymes can act as replacement for inactive endogenous enzymes (Winkel et al., 2004). CI-M6PR present on plasma membranes of most cells (Sly et al., 1978) is responsible for lysosomal targeting (Gary-Bobo et al., 2007). This receptor binds to mannose-6-phosphate (M6P) residues on lysosomal enzymes. For instance, ERT in form of exogenous β Glucocerebrosidase enzyme containing M6P can bind to M6PR on cell surfaces and be

taken up by cell via endocytic pathways (Sly et al., 1978), in what is known as cross-correction.

Currently, ERT has been approved for the treatment of numerous lysosomal storage disorders including Gaucher, Pompe, Fabry disease, MPS I, II, IV, VI, and VII, and lysosomal acid lipase syndrome (Li, 2018; Staretz-Chacham et al., 2009). In fact, ERT can significantly the prevent development of clinical problems in Gaucher's disease allowing individuals to lead a relatively normal life (Grabowski, 2012).

ERT is a gold standard in the treatment of PD for which exogenous produced rhGAA is infused into patients using i.v. injections for achieving cross-correction in target tissues (Herzog et al., 2012). In 2004 several patients started ERT with rhGAA produced in rabbit milk (Winkel et al., 2004), however this method of production of GAA was beset with difficulties and led to the development of another generation of rhGAA in 2006 produced in CHO cells, currently known with the commercial name of Myozyme (Kishnani, Hwu, et al., 2006). The first clinical trials have proven great success of ERT with rhGAA in IOPD patients without clinical signs of respiratory failure. In this cohort, ERT drastically improved survival rates up to 99% at 18th month and up to 95% at 36 months of age, in addition ERT decreased by 50% the need for invasive ventilation at 36 month of age (van den Hout et al., 2000). Furthermore, a reduction of left ventricle size was observed, and general conditions of heart were stabilised after few months of treatment and motor skills and functional independence, otherwise rarely seen in PD, were observed (e.g., sitting, crawling, standing, walking). When a wider age-range and clinical conditions of IOPD patients were included in the study, more contradictory outcomes were observed i.e. 35-60% ventilator-free survival, 27-35% mortality, 30-40% ventilator-dependent (Chakrapani et al., 2010; van Gelder et al., 2015). Despite ERT allowing motor-skills acquisition up to 4 years of treatment, only a small proportion of patients were able to maintain those improvements (van Gelder et al., 2015).

In LOPD trials using ERT showed that 38.5% patients with invasive ventilation support improved after 36 months of treatment, while 53.8% patients achieved stabilisation and 7.7% declined in conditions (Toscano & Schoser, 2013). In contrast, when treated patients were without invasive ventilator support 64.1% shown improvement, 32.1%

stabilized and 3.8% decline (Toscano & Schoser, 2013). In addition, LOPD patient treated with ERT for 36 months improved their forced vital capacity (FVC) in 51.6% participants, while in 13.7% patients the FVC was stabilized and in 34.7% the FVC declined. In addition to ventilation and FVC capacity, the motor performance was assessed by six-minute walk test (6MWT). In this walking test, 77.9% participants shown improvement, 8.2% were stabilized and 13.9% declined their walking ability (Toscano & Schoser, 2013). The overall improvement in LOPD on ERT was noticed and extended up to 4 years after starting the ERT regime, however with some cases showing an initial improvement followed by a decline of their overall conditions (Toscano & Schoser, 2013).

Current ERT routine consists of i.v. infusion of rhGAA fortnightly at a suggested dose of 20 mg/kg but it can be increased up to 40 mg/kg if patients are unresponsive to basal dose (Ronzitti et al., 2019). Currently, a second generation of rhGAA, with a higher affinity to CI-M6PR, is under evaluation in phase III clinical trial for LOPD (avalglucosidase alfa, Neo-GAA; Sanofi Genzyme) (Pena et al., 2019). In addition, another experimental rhGAA, with a higher content of M6P glycan residues, has being tested in a clinical trial in association with pharmacological chaperones (ATB200; Amicus Therapeutics). In preclinical studies, GAA enzymes enriched in M6P residues improved muscle function in *Gaa -/-* mice either alone or in combination with chaperones when compared to first-generation rhGAA (Macauley, 2016).

In spite of initial alleviation of clinical symptoms seen in PD patients treated with ERT including improved overall conditions in comparison to non-treated patients, and becoming independent ambulation (Chakrapani et al., 2010) with a stabilisation of respiratory functions (Schoser et al., 2017), following several long term studies patients' conditions eventually decline in time leaving them respiratory dependent and bedbound (Toscano & Schoser, 2013). In addition to the decline in PD patients overtime, ERT also presents with host humoral responses to exogenous enzyme diminishing the outcomes of the treatment (Do et al., 2019) (further detailed in section 1.3.6.2).

### 1.3.5.1. Cross-correction and CI-M6PR

Cross-correction is a naturally occurring secretion-reuptake process in which secreted proteins are transferred into adjacent cells (Fratantoni et al., 1968). This process also applies to lysosomal hydrolases. The trafficking of lysosomal enzymes to lysosomes is facilitated via M6PR (Braulke & Bonifacino, 2009; Fratantoni et al., 1968). 90% of lysosomal enzymes end up within lysosomal vesicles, however a small fraction escapes from lysosomal trafficking and are secreted (Braulke & Bonifacino, 2009). Fully functional secreted lysosomal hydrolases can be taken up from extracellular space by the cell from which it escaped or by neighbouring cells (Figure 1.8) through CI-M6PR present on cell membrane and enter endocytic pathway (Braulke & Bonifacino, 2009). ERT, to treat inter alia PD, utilises CI-M6PR to provide exogenous protein to enzyme-deficient cells through cross-correction mechanism.
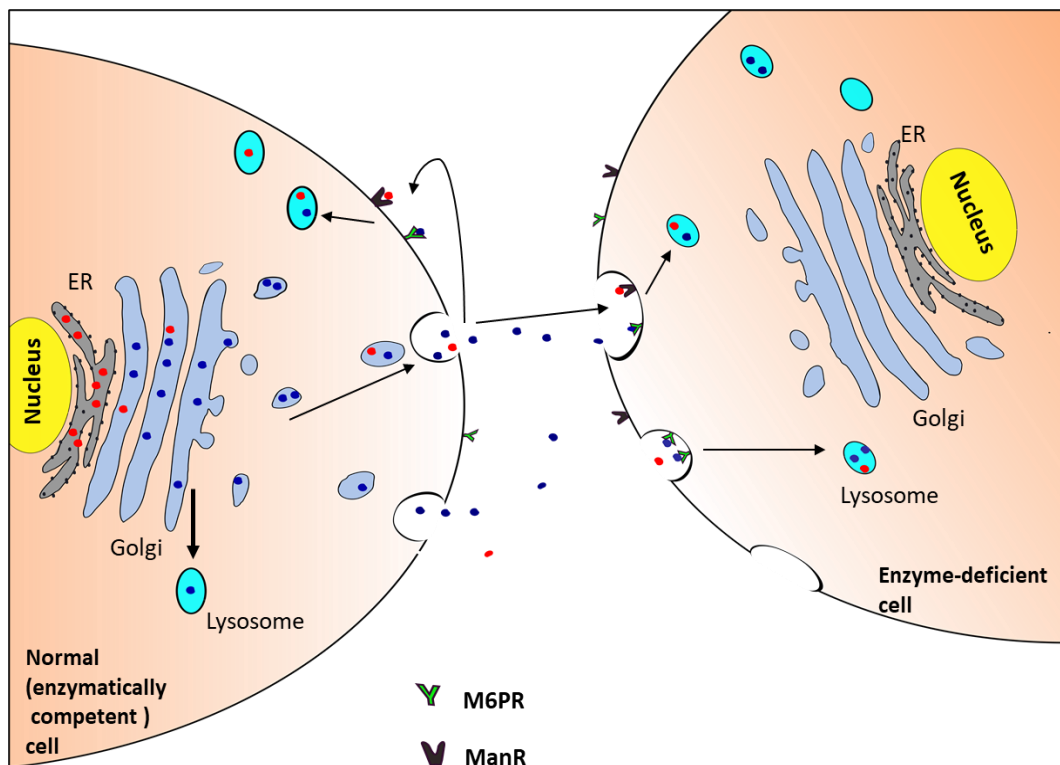


***Figure 1.8. Cross correction mechanism***. *Endoplasmic reticulum (ER) is the cellular structure where the glycosylation of newly formed lysosomal hydrolases (red circles) takes place. The addition of mannose-6-phospate residues (blue circles) follows in the Golgi apparatus, then the interaction with mannose-6-phospate receptor (M6PR) directs*

*their sorting into lysosomes (Lys). A small amount of lysosomal hydrolases is secreted from cells and consequently they can be taken up by neighbour cells through membrane bound CI-M6PR, if hydrolases possess phosphorylation on mannose residues, or mannose receptor (ManR), if the phosphorylation is not present. Both receptors lead to the activation of the endocytic pathway which leads to mature lysosomes. This process is called cross-correction and it can be utilised to deliver an exogenous lysosomal hydrolase into enzyme deficient cell.*

However, the efficacy and efficiency of ERT is limited by two factors: copiousness of CI-M6PR in different tissues and a binding affinity to its ligand M6P residues located on the protein surface (Ghosh et al., 2003).

CI-M6PR is 300kDa type I transmembrane glycoprotein containing two high affinity M6P binding sites in segment 3 and 9 and a solitary IGFII binding domain in segment 11 and 13 (Braulke & Bonifacino, 2009; van Meel & Klumperman, 2008) (Figure 1.9). Activation of CI-M6PR by its ligand leads to capture of the receptor-hydrolases complex into clathrin-coated vesicles, which fuse with early endosomes and later with lysosomes delivering lysosomal hydrolases to their final destination (Braulke & Bonifacino, 2009). The CI-M6PR has been extensively studied and it has shown to interact with a variety of ligands in a M6P independent manner (Dahms et al., 2008). The ligands in question consist of insulin-like growth factor II (IGF-II) (Dahms et al., 1994), retinoic acid (Kang et al., 1997), urokinase-type plasminogen activator receptor (uPAR) (Nykjaer et al., 1998) and heparanase (Wood & Hulett, 2008). Extensive studies on the interaction of different ligands with CI-M6PR have indicated that IGF-II has a 10-fold higher affinity to its binding site than M6P ligand has for its binding site (Devi et al., 1998). This discovery opened up a new area of study and potential therapies for PD, particularly to overcome the limitation found by ERT in skeletal and smooth muscle clinical outcomes (Kan et al., 2014).
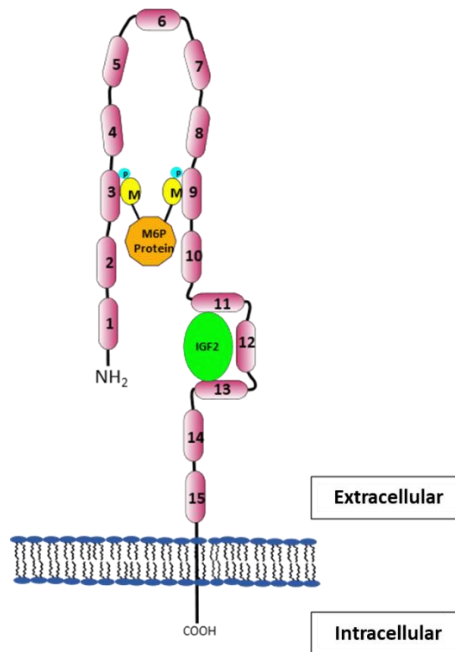
***Figure 1.9. Structure of cation independent Mannose-6-Phosphate Receptor (CI-M6PR).*** *CI-M6PR is a type I transmembrane glycoprotein receptor with molecular weight of 300kDA. It contains two M6P binding site in domain 3 and 9 and one IGF-II binding site in repeat 11 and 13.*

**1.3.5.2. Limitation of ERT in PD.**

Despite the fact that ERT is currently a gold standard in the treatment of LSDs, this method has several limitations. The success of ERT relies on the biodistribution of M6PR on plasma membrane of target cells. For instance, due to the insufficient expression of CI-M6PR on skeletal muscles (particularly relevant for PD) or osseous tissues the ERT therapeutic effect in these tissues is limited (Do et al., 2019). Another considerable drawback of ERT is the inability of rhGAA to penetrate and successfully cross the BBB and accomplish a cross-correction in the CNS, which becomes more important in light of the neurocognitive defects that are being seen in PD patients as they survive longer on ERT (Ebbink et al., 2018; van den Hout et al., 2000).

In addition, the infusion of exogenous protein causes, in various levels, the patient's own humoral immune response to generate antibodies to the exogenous enzyme thereby

neutralizing the infused enzyme and lessening treatment outcomes and the overall survival (Kishnani et al., 2010). This is particular related to ''CRIM negative'' patients who have no residual GAA protein expression, consequently no presence of GAA antigen and therefore of immune tolerance to rhGAA protein (Desai et al., 2019). CRIM negative patients have a very poor clinical prognosis and have a higher dependency on mechanical ventilation. In addition, ERT infusion require often hospitalisation with many hours needed for the administration of ERT (Desai et al., 2019).

### 1.3.5.3. CI-M6PR alternatives - Muscle homing peptide (MHP)

Despite all the efforts to increase the affinity of rhGAA binding to CI-M6PR in order to mitigate clinical symptoms of PD, in particular in muscle tissues, the results are still far from satisfying (Baik et al., 2021). Therefore, alternatives to M6PR/M6P binding as entry point are investigated to target skeletal muscles (Gao et al., 2014; Samoylova & Smith, 1999). A method that uses M13 bacteriophages coding for random peptide libraries on their surface proteins and biopanning in C2C12 murine myoblast has been employed to identify muscle homing peptides (Barry et al., 1996; Smith, 1985). Out of all bacteriophages used those carrying a Muscle Specific Peptide (MSP), also known as ASSLNIA, on their surface were isolated and their biodistribution in mice showed 5-fold and 2-fold increase of phage recovery in skeletal muscle and heart, respectively, when compared to wild type bacteriophages without insert (Samoylova & Smith, 1999). In another study conducted by Gao *et al.* a different type of phage Ph.D-12 was used to express an array of various short peptides on its surface. One of the peptides , 12 amino acids peptide, known as M12 or RRQPPRSISSHP, showed very promising muscle targeting ability in dystrophic mice (Gao et al., 2014). Indeed, M12 associated to morpholino showed enhanced therapeutic effect in skeletal muscles of dystrophic mice compared to MSP associated or unmodified morpholino (Gao et al., 2014). Although the use of MHPs has great potentials to target skeletal muscles (Gao et al., 2014; Samoylova & Smith, 1999), the cellular counterparts and the entry points remain yet unknown.

### 1.3.6. Hematopoietic stem cell transplant

With the advancement of medicine and technologies hematopoietic stem cell transplants-(HSCT) are no longer limited to treat immune and blood disorders but can be also used to treat successfully some forms of LSDs (Lamsfus-Calle et al., 2020; Thomas et al., 1977; Thomas et al., 1957).

Allogeneic HSCT was in clinics before ERT for the treatment of MPS-I (Hobbs et al., 1981; Thomas et al., 1957). The rationale behind HSCT was that healthy donor's stem cells can supply the lacking enzyme and also that the progeny of HSCs can transmigrate to visceral organs to correct disease (Krause et al., 2001; Priller et al., 2001). Enzymes secreted by progeny of allogeneic HSCs can be taken up by affected organs in the process called cross-correction (Figure 1.8), correcting the phenotype of the disease (Sands & Davidson, 2006). In addition, donor's stem cells have been shown to migrate in CNS and differentiate, to a degree, in microglia cells (Asheuer et al., 2004) with the ability to deliver the missing enzyme to neurons (Biffi et al., 2006). Currently, HSCT is the first line treatment for Hurler disease (Hobbs et al., 1981). HSCT conducted within the first 9 months of age can completely mitigate neurological symptoms, but transplant occurring after 26 months of age can only permit peripheral improvement but do not avert cognitive decline (Poe et al., 2014). In addition to Hurler disease, patients with Krabbe disease also benefit from HSCT although outcomes are more limited (Langan et al., 2016).

### 1.3.6.1. Hematopoietic stem cells

Hematopoietic stem cells (HSCs) are self-renewing multipotent stem cells that in adult organisms reside in bone marrow and give rise to specific haematopoietic lineages including myeloid and lymphoid progenitors and in consequence all blood lineages (Ng & Alexander, 2017).

It is hypothesised that the most primitive pluripotent stem cells (PSCs), which are the most naïve and the least differentiated, gives rise to long-term repopulating HSCs (these cells still maintain the self-renewal ability), short-term repopulating HSCs (they have lost the ability to self-renewal but can still differentiate into all blood lineages), endothelial

progenitor cells (EPCs) and mesenchymal stem cells (MSC) (Ratajczak, 2008) (Figure 1.10). Mature HSCs can differentiate into common myeloid progenitors (CMP), which can give rise to macrophages, granulocytes, platelets, erythrocytes, and common lymphoid progenitors (CLP), which eventually give rise to lymphocytes T and B as well as NK cells and dendritic cells (Bryder et al., 2006)
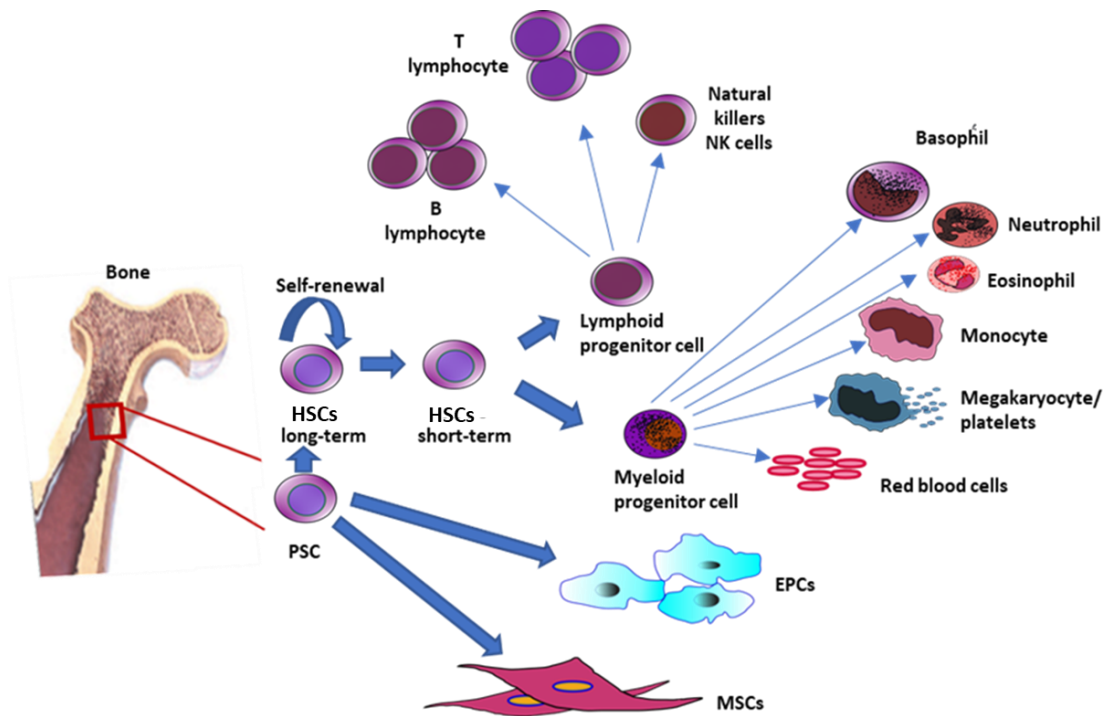


***Figure 1.10. Origin and hierarchy tree of hematopoietic stem cells and their progeny.*** *Pluripotent stem cells (PSC) originating in bone marrow give rise to long-term hematopoietic stem cells (HSCs) possessing a self-renewal ability to sustain constant pool of themselves and newly form short-term HSCs. In turn, short-term HSCs give rise to all blood lineages. It is also hypothesised that PSC give rise to endothelial stem cells (EPCs) and mesenchymal stem cells (MSCs).*

### 1.3.6.2. Limitation of hematopoietic cell transplantation

Although HSC transplant is a cell therapy with established techniques and history of clinical success, it poses challenges in finding suitable cell donors and in the

management post-transplant side effects, such as graft versus host disease (Sands & Davidson, 2006). Allogeneic HSC transplant as a therapeutic treatment in MPSI or Hunter syndrome can lead to graft vs. host in patients who then need to be under long term immunosuppression (Biffi, 2017). In addition, in LSDs the supply of enzyme from progeny of transplanted HSCs in poorly accessible tissues i.e. bone or CNS is still far from optimal under physiological expression of the enzyme in HSCs (Malgieri et al., 2010).

However, these limitations can potentially be overcome by the use of autologous HSCs and gene therapy. HSCs represent a great gene transfer platform due firstly to their self-renewing ability, secondly their ability to differentiate in multiple cell lineages including tissue resident macrophages, which can populate tissues/organs among which is the CNS, and thirdly their ability to induce immune tolerance towards new expressed genetic materials (Bryder et al., 2006). Through engineering of the number of gene copies in the modified HSCs or through the use of appropriate promoters, it is possible to overexpress the enzyme from gene modified cells and thereby deliver more enzyme to target tissues. Once *ex vivo* gene transfer is undertaken and the gene engineered HSCs are reintroduced back to the organism they came from, they can persistently express the gene of interest across all systems potentially alleviating clinical manifestations of treated diseases including LSDs (Daniel-Moreno et al., 2019).

## 1.4. Gene Therapy for LSDs

The goal of gene therapy in LSDs is to supply constant therapeutic levels of active enzymes and/or lysosomal membrane proteins to ease both CNS and systemic symptoms (Gritti, 2011). The transgene encoding a functional enzyme is provided to cells via an inactive form of viral vector like lentivirus (LV), or adeno-associated virus (AAV) (Aronovich et al., 2017; Sands & Davidson, 2006). Two major delivery methods, *in vivo* and *ex vivo* (Figure 1.11), are used in gene therapy and both with the aim of correcting LSD phenotypes (Amalfitano et al., 1999; Kyosen et al., 2010).
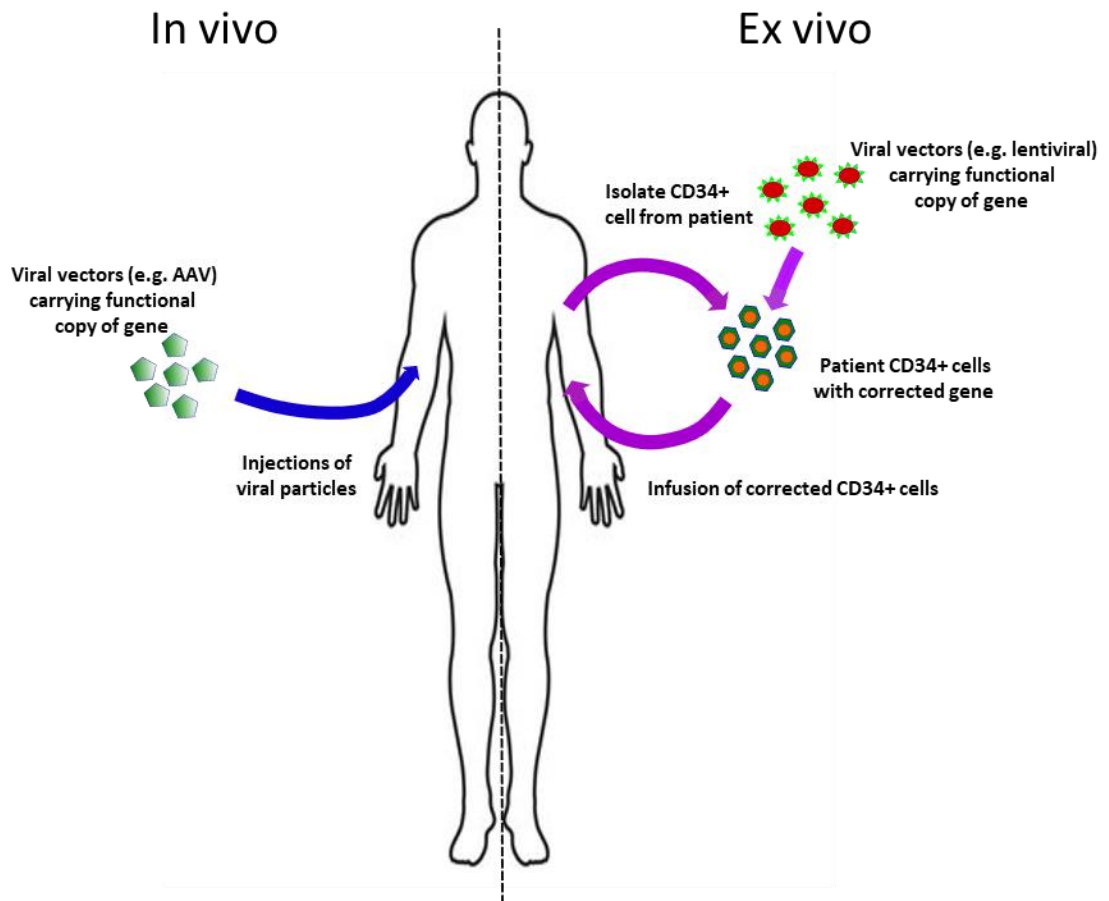
*Figure 1.11. Gene therapy models.* *There are two principal ways to execute a gene therapy treatment. On the left, the delivery is through in vivo injection of viral vectors bearing a transgene. This approach is mainly used for the delivery of adeno-associated viruses (AAVs). On the right, the delivery is through ex vivo manipulation of autologous human CD34+ hematopoietic progenitor cells, which are harvested from a patient and transduced with lentiviral vectors in vitro before being re-introduced to the patient.*

### 1.4.1. Viral vectors used in LSDs gene therapy.

Currently there are two major types of vector platforms used for LSDs gene therapy: adeno-associated virus (AAV) varying from type 1 to 9 (Gao et al., 2005; Samulski & Muzyczka, 2014) and also HIV-based lentivirus (Escors & Breckpot, 2010; Piras et al., 2020). Each of the platforms bring specific advantages and disadvantages in gene therapy and are described in detail below.

## 1.4.1.1. AAV structure

AAVs belong to the family of *Parvoviridae*, small 25nm nonenveloped viruses with icosahedral symmetry (Pennazio, 2007; Rey, 2007) (Figure 1.12 A), containing about 5kb long single stranded DNA (Srivastava et al., 1983). AAVs were first discovered in 1965 as a contamination of adeno virus purification (Atchison et al., 1965). AAVs have a defective replication system and for successful replication requires a presence of a helper virus such as herpesvirus or adenovirus (Chiorini et al., 1997). AAV viral genome contains three open reading frames (orfs), which encode eight viral proteins expressed from three promoters (Figure 1.12 B) (Srivastava et al., 1983). However, the mature functional viral capsid is composed by three viral proteins encoded by only one orf (cap) and encloses the ssDNA genome (Sonntag et al., 2010). Inverted terminal repeats (ITRs) flank the coding region of AAV (McLaughlin et al., 1988). The ITRs are *cis*-active structures required for the production of recombinant AAV (rAAV) vector (Samulski et al., 1989). AAVs upon the entry to cells binds to surface sugars present on proteoglycans including galactose, heparin sulphate or sialic acid and to cellular plasma membrane receptors such as integrin or fibroblast growth factor receptor (Nonnenmacher & Weber, 2012). The stimulation of cellular receptors activates endosomal uptake by clathrin-coated vesicles and also clathrin-independent carriers/GPI-enhanced (Agbandje-McKenna & Kuhn, 2011). The contact with acidic pH of endosomes is crucial for AAV structural changes (Sonntag et al., 2006) which allow the rupture of endosomes and the entry into the cytoplasm of the host cell (Girod et al., 2002). *In vitro* genomic studies have indicated that rAAV DNA molecules are converted into circular structures and persist largely as episomal genetic material containing copies of transgene cassette, and integration into the DNA host occurs at very low frequency (Nakai et al., 2001). Up to date 10 serotypes of AAVs are defined and used for gene transfers in many diseases and in many hosts with distinct tissue tropisms (Gao et al., 2005). rAAV vectors are devoid of any harm-causing viral gene structures that could cause unwanted immune response, however studies have shown that re-administration of rAAV into circulation induce anti-capsid rAAV host humoral response damping down therapeutic effect (Zaiss & Muruve, 2005). In addition, another disadvantage of rAAV might be a fact that their gene transfer forms

episomal concatemers that in dividing cells is lost through the dilution effect of cell division, since the episomal DNA does not follow the replication cycle alongside the DNA of host cell (Gao et al., 2005; Nakai et al., 2001). Nevertheless, rAAVs due to their variety and diversity in tissue tropism (Passini et al., 2003) represent an attractive vehicle for *in vivo* gene therapy in LSDs and other diseases (Sands & Davidson, 2006).
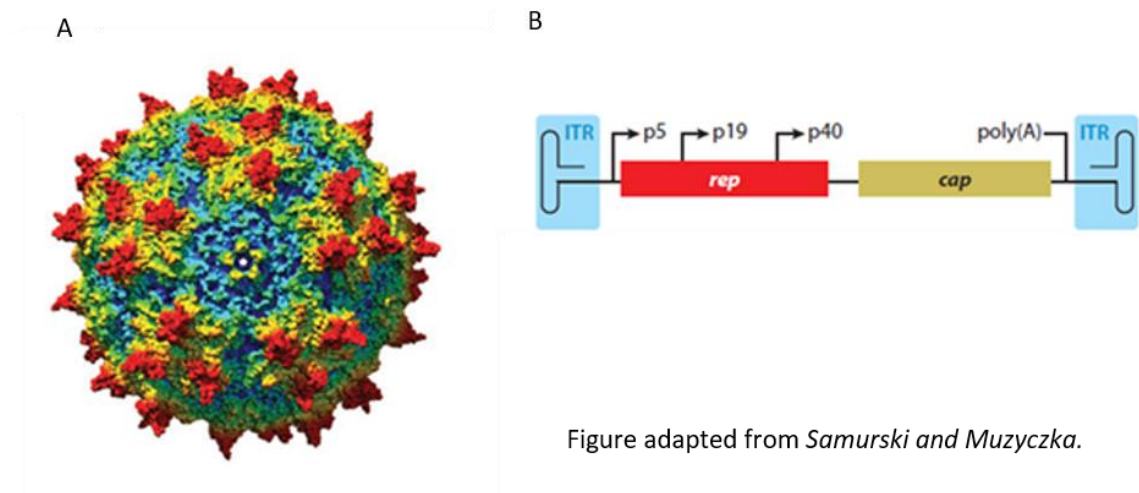


Figure adapted from *Samurski and Muzyczka.*

*Figure 1.12. Structure of AAV. A) Cristal structure of AAV capsid particle. B) Map of AAV genome with rep containing three open reading frames (orfs) that code for all functional proteins and cap fragment containg genes encoding capsid protein components. Figure adapted from Samulski and Muzyczka (Samulski & Muzyczka, 2014).*

### 1.4.1.2. HIV-based Lentiviral vectors

Lentiviruses (LVs) belong to the family of *Retroviridae* (Retroviruses), contain single strand RNA, which upon entry to host cells, is retrotranscribed into DNA and integrated into the genome of the host (Wilusz, 2005). Retroviruses are spherical viruses between 80-120nm in diameter (Vogt & Simon, 1999). LVs have been engineered using various agents including feline immunodeficiency virus, equine infectious anaemia virus and human immunodeficiency virus (HIV-1) (Azzouz & Mazarakis, 2004; Kay et al., 2001). HIV-1 genome contains 9-10kb and comprises of several non-coding fragments, which control gene expression and synthesis of essential protein, followed by sequences encoding for regulatory and accessory proteins in addition to enzymatic and structural

genes, env, gag, and pol, (Figure 1.13 B) that are crucial genes for the correct assembly of the viral particle and interaction with cellular receptors on the host cell (Katz & Skalka, 1994; Pluta & Kacprzak, 2009). Viral genome also includes Long Terminal Repeats (LTRs) containing promoter/enhancer fragment allowing vector's gene expression (Pluta & Kacprzak, 2009). In HIV-based LV vectors, the expression cassette is incorporated between the LTRs and contains the gene of interest and an internal promoter is used to drive the transgene expression (Pluta & Kacprzak, 2009). The expression cassette can hold up to 9kb providing an excellent platform for gene transfer (Milone & O'Doherty, 2018). Currently used HIV based vectors are self-inactivating vectors in which the viral enhancers and promoters in U3 region of LTR 3' are removed for safety (Lutzko et al., 2003). The deletion of viral genes prevents viruses from uncontrolled replication upon entry to the host cells (Pluta & Kacprzak, 2009). The tissues specific tropism of LV is achieved by removal of the native envelope proteins and replacing them with substitute envelope glycoproteins, most often glycoproteins from vesicular stomatitis virus (VSV-G), provided *in trans* during *in vitro* viral assembly (Wong et al., 2004). This manipulation of envelope glycoproteins allows pseudotyping and in particular the VSV-G pseudotyping gives broader tissue tropism and improves viral particles stability during production and concentration (Kang et al., 2002). Due to LV loading capacity and integration property, HIV-based LV vectors became a promising tools for *ex vivo* gene transfer in many lysosomal storage diseases including MPSI, MPSII, MPSIIIA, Niemann-Pick type A and B (NPA, NPB) and MPS-VII (Fraldi et al., 2018; Sevin & Deiva, 2021; Wada et al., 2020; Sands & Davidson, 2006; Stein et al., 2001).
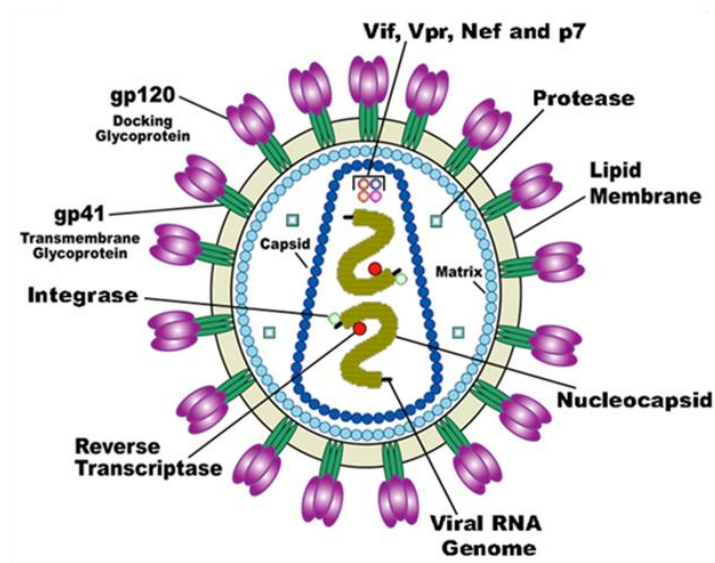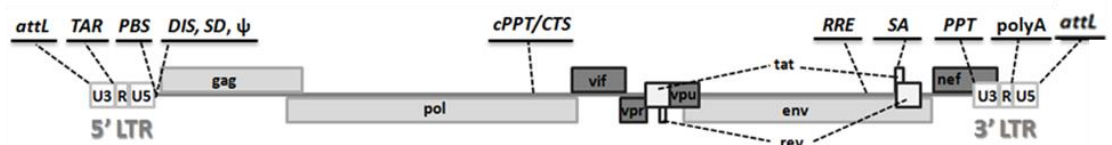
Figures adapted from a) *O'Keefer et al.* and b) *Tomás et al.*

***Figure 1.13. LV structure. A)*** *Structure of a lentiviral particle showing transmembrane glycoproteins, responsible for the receptor binding and viral entry into the host cells, immerse in the viral membrane (lipid layer) and the viral core containing the RNA genomic material, reverse transcriptase. Accessory proteins are also annotated (Nef, Vif, Vpr, and p7) that are not required for viral replication, nonetheless helping in the proficiency of this process. **B)** Map of genetic components in lentivirus genome including gag, pol and env which code for structural viral protein and viral reverse transcription polymerase. Those are flanked by 5' and 3' long terminal repeat (LTR) containing both attL left and right attR attachment sites, U5' and U3' untranslated region and R5' and R3'repeat elements. Furthermore, on 5' LTR is present TAR, transactivation response element, followed by PBS region responsible for primer binding site, dimerization signal*

*(DIS), splice donor site (SD) and finally ψ packaging signal. Central polypurine tract (cPPT) and central termination sequence (CTS) are located within pol sequence. Remain Reverse response element (RRE), splice acceptor site (SA), polypurine tract (PPT), polyadenylation signal (polyA) to complete lentiviral genome. Figures adapted from A) O'Keefer et al. (O'Keefer et al. 2015) B) Tomás et al. (Thomas et al. 2013)*

### 1.4.1.3. Safety measures and the critical role of transgene promoters

Vectors used in gene therapy have been stripped of majority of their wild type genome in order to prevent these vectors from activating oncogenes if incorporated in the host genome (Bushman, 2007) and also to abolish uncontrolled replication in host cells (Lutzko et al., 2003; Milone & O'Doherty, 2018). Viral promoters, for safety reasons, have been removed from U3 LTR 3' region in LV vectors and replaced by internal cellular promoters derived from human genes including ubiquitously expressed elongation factor-1α (EF1α) or phosphoglycerate kinase (PGK),  (McBurney et al., 1991; Montiel-Equihua et al., 2012; Sawado et al., 2003) or human lineage specific promoters. Also, the gammaretroviral derived promoter MND has been used to drive expression of transgene in viral vectors (Lutzko et al., 2003) and in several contexts regulatory element like β-globin locus control region (LCR) have been crucial for the overexpression of the transgene specifically in red blood cells. These promoter/regulatory elements when coupled with a transgene in a transfer cassette permit a persistent expression of the transgene (Montini et al., 2006; Schambach et al., 2006). In particular, the use of EF1α and PGK promoters is proven to produce persistence transgene expression in transduced cells providing high level of therapeutic effect (Montiel-Equihua et al., 2012; Sawado et al., 2003). However, a careful consideration of promoters/ regulatory elements preceding the transgene needs to be undertaken to meet the gene expression needed for therapeutic efficacy as each promoter might have a specific tissue/ cell lineage preference i.e. LCR fragment enhances transgene expression in red blood cell lineages or EF1α promoter might be more suitable than PGK promoter when considering liver cells as a gene therapy target etc. (Montiel-Equihua et al., 2012; Piras et al., 2020; Sawado et al., 2003).

### 1.4.2. *In vivo* gene therapy

*In vivo* gene therapy refers to direct injection into tissues or in blood circulation of a gene transfer vector either LV or AAV. Vectors, upon the entry into host cells, either integrate their genetic material into the DNA of the host for instance through stable long-terminal repeats (LTRs) (Gray et al., 2010) or remain in cytoplasm as a transient episomal vector without fusion with the DNA of the host, in the case of AAV vectors (Nakai et al., 2001). The outcome of gene therapy in LSDs largely depends on the nature of the viral vector used: AAV1-9 serotypes or different envelope glycoproteins used for LV pseudotyping can alter vector performance, stability and tissue tropism (Cronin et al., 2005; Zincarelli et al., 2008).

AAV vectors were used in intramuscular injection in Pompe, Fabry and MPS-VII disease, however this approach showed high level of enzyme production only local to the injection site in contrast to undetectable level of enzymes in circulation (Fraites et al., 2002; Takahashi et al., 2002; Watson et al., 1998). On the contrary, the intravenous injections of AAV vectors targeting liver, in animal studies of Fabry, MPSVII and Pompe disease, resulted in production and secretion, into the circulation, of lysosomal enzymes by hepatocytes in high therapeutic levels, leading to permanent mitigation of symptoms of aforementioned diseases (Sun et al., 2005; Takahashi et al., 2002; Watson et al., 1998). Intracerebral injections of AAV or LV vectors were performed in clinical studies in MPS-IIIA (101), MPS-IIIB (Di Domenico et al., 2009; Sawamoto et al., 2018), MLD (Consiglio et al., 2001), Niemann-Pick type A (Passini et al., 2005) and globoid cell leukodystrophy (GLD)(Lin et al., 2005) showing mitigation of neurological symptoms and stable expression of lysosomal enzymes (Hocquemiller et al., 2016). LV vectors were also being used for *in vivo* gene therapy in MPS-I and MPS-VII showing constant high-level expression of enzyme followed by reduction in the build-up of lysosomal materials in numerous tissues (Kobayashi et al., 2005; Stein et al., 2001).

In the past decade several different AAV serotypes were tested in animal models of PD ranging from AAV1 (Mah et al., 2010), through AAV2, 5, 6, 8 (Salabarria et al., 2020) to AAV 9 (Doerfler et al., 2016) and targeting the liver (Kishnani & Koeberl, 2019), muscles (Puertollano & Raben, 2018) or CNS (Byrne et al., 2019). In addition, intramuscular injections of AAV9 vectors containing GAA transgene in a mouse model of PD showed

total correction of glycogen storage, however the improvement was only seen in the area local to the injections and this method of delivery required the co-administration of immunosuppressive drugs to block humoral immune responses (Mingozzi & High, 2013). Several preclinical studies of systemic delivery of AAV vectors containing muscle expression cassette for the GAA transgene showed significant glycogen clearance in muscles, greatly improving respiratory and cardiac functions (Puzzo et al., 2017; Salabarria et al., 2020). However, targeting muscle through systemic injections of AAV vectors requires enormous doses of vector, higher than 10e14 vector genomes/Kg (Fraites et al., 2002; Ronzitti et al., 2019), which in LOPD is particularly difficult to achieve. As valuable alternative, liver targeting strategy has been developed to be administrated by an intravenous injection of AAV with GAA expression cassettes in PD patients (Kishnani & Koeberl, 2019). Liver is a key biosynthetic and secretory organ of the body, secreting a great variety of proteins and enzymes into the bloodstream and therefore liver has a great potential to become a "factory" of transgene enzymes (Kishnani & Koeberl, 2019). Systemic administration to target the liver was employed in mouse study to assess safety and efficiency of AAV vectors bearing GAA expression cassette with a great success achieving cross-correction, reduction in total percentage of glycogen accumulation and rescuing of muscle and cardiac tissues in *Gaa-/-* mice (Puzzo et al., 2017). An additional benefit from targeting liver with AAV gene therapy is the induction of antigen-specific tolerance towards proteins expressed and secreted by hepatocytes (Kattenhorn et al., 2016). This tolerance is partly due to TGF-β-dependent induction of CD4+ CD25+ Foxp3+ regulatory T cells (Treg) that readily suppress T and B cells resulting in the suppression of humoral response against GAA in *Gaa-/-* mice study (LoDuca et al., 2009; Puzzo et al., 2017). However, the GAA produced in the liver and secreted into the bloodstream will still not exhibit the ability to cross BBB (Zhang et al., 2018). Therefore, intraspinal injection of AAV gene transfer carrying GAA transgene cassette was explored in Pompe animal model resulting in normalized spinal cord GAA activity and decrease in spinal PAS staining 1 month post injection  (Qiu et al., 2012).

However, *in vivo* gene therapy has several limitations. Current use of AAVs in gene therapy suffers from episomal transgene expression (Nakai et al., 2001), with majority

of AAV vectors not able to integrate their genes into the host genome leading to a decrease in number of transgene copies with each cell division. This reduction of copies of transgene in cells of growing organs i.e., liver in time limits the therapeutic effect. A second dose of AAVs, in a later stage of patient's life, is unfeasible due to humoral immune response which neutralise AAV vectors (Suhonen et al., 2006; Wong et al., 2006). Direct infusion of vectors to the brain or spinal fluid is an invasive method to treat CNS through *in vivo* gene therapy (Lim et al., 2019; Schuster et al., 2014). In addition, viral vectors do not radially diffuse from the injection site when delivered through intrathecal injections limiting therapeutic range and outcomes across CNS and consequently requiring several complicated and potentially life-threating intracranial injections (Macauley, 2016). New generation self-complementary (sc) AAV9 vector, administrated through i.v. injection, showed full correction of motor neuron in a model of spinal muscular atrophy without pharmacological disruption of BBB (Duque et al., 2009). The transgene carried by scAAV9 was present in cells of the spinal cord of treated mice proving that the vector was able to penetrate tight cellular junctions of BBB. However, the mechanisms through which the scAAV9 achieved penetration of spinal cord remain unknown (Duque et al., 2009). This study shed light on a non-invasive long-term correction of CNS pathologies seen in the majority of LSDs including PD without a surgical intervention.

### 1.4.3. *Ex vivo* gene therapy

*Ex vivo* treatment strategy is based on the isolation of progenitor stem cells from a patient, modification of those ex-vivo with lentiviral vectors containing an appropriate expression cassette and the correct version of the gene of interest, and then reintroduction of corrected cells into patient's organism and establishment of a stable graft (Blaese et al., 1995; Porteus, 2011). This method can employ variety of stem cells as a gene transfer platforms, ranging from HSCs, embryonic stem cells (ESCs) to induced pluripotent stem cells (iPS) (Porteus, 2011). Several gene therapy trials using *ex vivo* modification of patients' own HSCs are currently ongoing with CRISPR/Cas9, LV and AAV vectors (Aiuti et al., 2013; Daniel-Moreno et al., 2019; Lamsfus-Calle et al., 2020) (Figure 1.14).
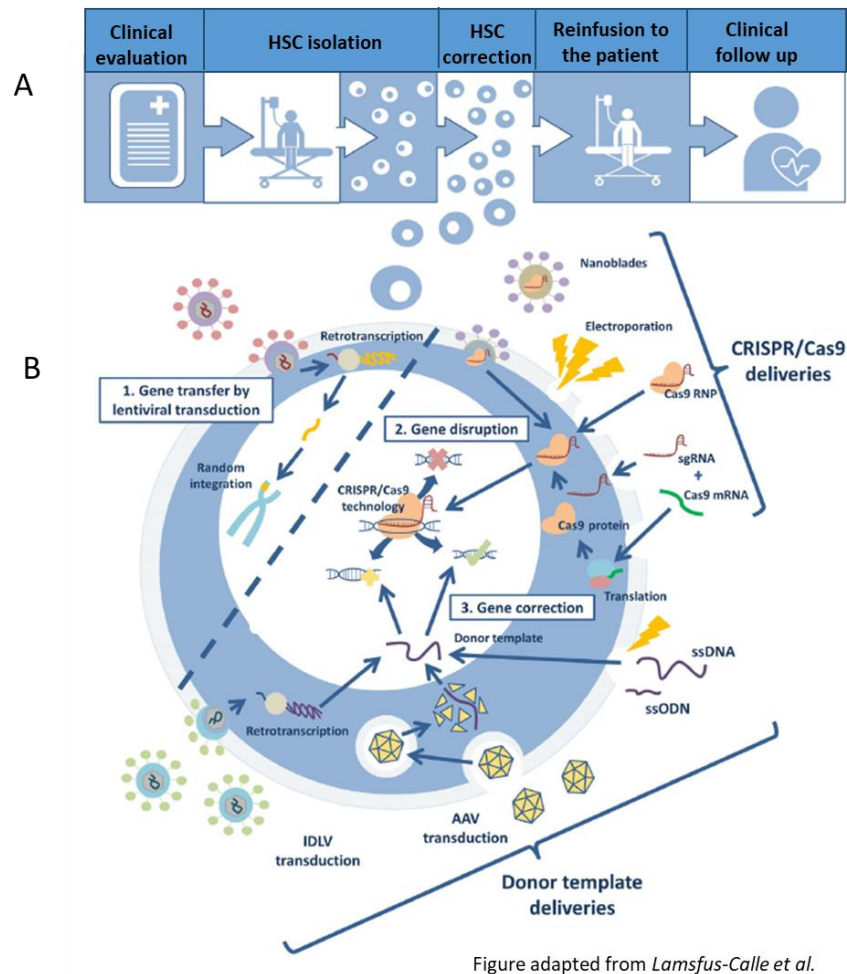
**Figure 1.14 Modality of ex vivo gene therapy employing hematopoietic stem cells. A)** *Autologous HSC transplant starts with the isolation of HSCs which have been mobilised from bone marrow into a blood circulation. Harvested HSCs then are genetically modified ex-vivo. Consequently, gene corrected/modified HSCs are infused back into the patient.* **B)** *Representation of the currently available ex vivo techniques in gene therapy: 1. LV vectors for transgene transfer, 2) gene disruption though CRISPR/Cas9 machinery followed by 3) gene correction by homologous repair with a donor templated provided either through AAVs or as single strand nucleic acid. Figure adapted from Lamsfus-Calle et al. (Lamsfus-Calle et al., 2020).*

LV vectors due to their biology became a main gene transfer vector in *ex vivo* gene therapy (Ronzitti et al., 2019). Current LV vectors have the capacity to transduce both dividing and non-dividing cells and therefore can readily transfer genes into quiescent

stem cells (Ronzitti et al., 2019). In addition, LVs incorporate their genetic material permanently into DNA of host cells making the transferred gene enduringly present and persistently expressed (Milone & O'Doherty, 2018). Genetically modified stem cells and their progeny provide a constant source of enzyme secreted into blood circulation, which in turn can cross-correct the systemic pathology of LSDs (Sands & Davidson, 2006). In particular, CNS can benefit from *ex vivo* gene therapy as transplanted HSCs can migrate across the BBB into the CNS and differentiate in microglia like cells, which can, in turn, provide neurons with deficient enzyme (Staba et al., 2004; West et al., 2009), with the advantage that autologous stem cell transplants do not cause graft vs host immune response. In addition, patients treated with ex-vivo gene therapy using autologous HSCs develop immune tolerance towards the transgene which now encodes a self-recognise protein (Biffi, 2017), so it can be easily tolerated even by CRIM negative patients (Desai et al., 2019).

In a clinical trial conducted by Biffi *et al.* (Biffi et al., 2013) to treat metachromatic leukodystrophy (MLD), LV vectors were used to transfer *ex vivo* the gene coding for arylsulfatase A (ARSA) in human HSCs (Biffi et al., 2013). After the reintroduction of genetically modified patient's HSCs, significant enzymatic activity of ARSA was present in all lineages derived from transplanted HSCs and also in cerebrospinal fluid (Biffi et al., 2013). Furthermore, clinically relevant improvements were also observed in clinical trials of *ex vivo* HSCs transplants for MPS-I (Fraldi et al., 2018).

In addition, recent animal studies using murine hematopoietic stem cells and LV vectors with GAA transgene cassette proved that *Gaa-/-* animals after infusion of genetically modified HSCs have shown GAA activity in affected tissues and reduction of glycogen when compared with untreated *Gaa-/-* animals (Piras et al., 2020; van Til et al., 2010). These studies are a proof of concept regarding the use of genetically modified HSCs with designed hGAA carried by LV vectors which are capable of permanently reversing the pathology of PD in a mouse model.

### 1.4.4. Gene editing

### 1.4.4.1. CRISPR/Cas9 technology

CRISPR/Cas9 (clustered regularly interspaced short palindromic repeats/caspase 9) is a new technique of gene editing with the potential to correct specific genetic mutations by insertion of a functional copy of a defective gene directly in the genome of a host cell (Charpentier & Marraffini, 2014; Daniel-Moreno et al., 2019; Hille & Charpentier, 2016; Platt et al., 2018). This approach is based on the use of a short RNA sequence to direct the endonuclease Cas9 protein to a specific site in the host cell genome where it can cut double stranded DNA (Hille et al., 2018; D. Wang & Gao, 2014). The double strand breaks in DNA can be repaired through an error prone process known as non-homologous end joining; however, if a template with homology arms is provided the cleavage can be repaired through homology directed repair (Yin et al., 2017). This last mechanism can allow the replacement of a faulty/mutated part of a gene with the correct sequence (Christensen & Choy, 2017; Doudna & Charpentier, 2014). Templates are generally single strand DNA fragments, that once incorporated into the cleaved DNA by homologous recombination, serve as template for the synthesis of the complementary DNA strand resulting in the restoration of native gene sequence in a process called control gene alteration (Christensen & Choy, 2017; Daniel-Moreno et al., 2019). MPS-I and II mouse model studies evaluated safety and long-lasting benefits of gene editing using CRISPR/Cas9 approach providing an alternative method for the treatment of some LSDs (Poletto et al., 2020). In addition, *in vitro* study by Lu *et al.* (Lu et al., 2003) on the correction of mutations in *GAA* gene by modification of single-stranded oligonucleotides has rescued the correct genotype of GAA providing a foundation for gene editing using CRISPR/Cas9 technology and possible *in vivo* animal models of PD.

### 1.5 *In vitro* and *in vivo* PD models.

*In vitro* models of Pompe Disease are very useful tools to address fundamental questions regarding lysosomal glycogen accumulation and autophagic debris and test different treatment regimen. However, isolation and culture *in vitro* of muscle tissue/ myoblasts

of PD patients to study the disease pathogenesis/ pathology is very challenging and potentially unsustainable. Instead, immortalised cell lines and murine models of the disease are used to interrogate potential therapeutic approaches.

There are several commercially available murine cell lines of immortalised muscle or myoblast-muscle stem cells, which are capable of developing structures of multinucleated myotubes resembling the simplistic structure of skeletal muscle. One of the most widely used *in vitro* muscle myoblast cells are C2C12 cell lines that have been subcloned from a myoblast line established from adult C3H mouse leg muscle (Yaffe & Saxel, 1977).  Another muscle cell line used in vitro is H2K 2B4, an immortalised satellite cell-derived cell-line (Muses et al., 2011). In addition to murine models of skeletal muscle cells there are also human cell lines available commercially i.e., Human Skeletal Myoblasts (HSkM) and immortalized myogenic human cell lines TERT/cdk4 (Cheng et al., 2014; Thorley et al., 2016) .

However, all aforementioned cell lines do not fully represent Pompe pathology seen in patients as these cell lines are free of mutations in *GAA* gene and therefore free of lysosomal glycogen accumulations. Nevertheless, generation of a GAA knockout cell line using gene editing technology is a readily achievable goal with current development of molecular tools. In this study, I used CRISPR Cas9 methodology to successfully generate C2C12 *Gaa-/-* myotubes.

In addition to *in vitro* immortalised myoblast cell lines, *in vivo* mouse models are readily available. There are several Pompe disease mice models available generated by gene editing mice embryos. The most common preclinical model of Pompe disease is the Raben model of GAA knockout mice (B6;129-Gaa$^{tm1Rabn}$/J) that demonstrates pathogenesis of PD with progressive muscle glycogen storage over time and progressive muscle weakness (Raben et al., 1998). This model was generated by inserting into *Gaa* exon 6, a neomycin resistance cassette leading to complete cessation of the GAA protein production. In addition, Huang *et al.* attempted, using CRISPR Cas9 to generate an exact mutation that had been observed in Pompe patients c.1826dupA (p.Y609*) (Hermans et al., 2004) leading to generation of *Gaa*$^{c.1826dupA}$ murine model (Huang et al., 2020). This

model is believed to better pathology seen in IOPD patients i.e., muscle damage due to glycogen accumulation, hypertrophic cardiomyopathy, and respiratory dysfunction compared to Raben's model (Huang et al., 2020).

The pathogenesis of PD starts with progressive lysosomal glycogen accumulation and build-up of autophagic debris leading to the classic pathology of PD which is degeneration of skeletal and cardiac muscle in 3-month-old *Gaa-/-* mice vs. wild type (WT) mice; illustrated in Figure 1.15 A. Glycogen load was determined by periodic acid-Schiff (PAS) staining of heart, diaphragm, and gastrocnemius muscle. PAS staining indicates aggregates of glycogen and also muscle fibers in Gaa-/- vs WT look more irregular in structure with unusual spaces between muscle cells. In addition, echocardiography of *Gaa-/-* mice show hallmarks of hypertrophic cardiomyopathy with substantial increases in left ventricular posterior wall diameter (LVPWd) and intraventricular septal diameter (IVSd) (Figure 1.15 B)

Both assessments conducted on cellular/muscle level and heart echocardiography of *Gaa -/-* mice indicated a presence of typical hallmarks of PD seen in IOPD. This mirroring of IOPD/LOPD pathology in *Gaa-/-* mice models allow us to study pathogenesis of PD and explore novel forms/delivery routes of therapies for this devastating disease.
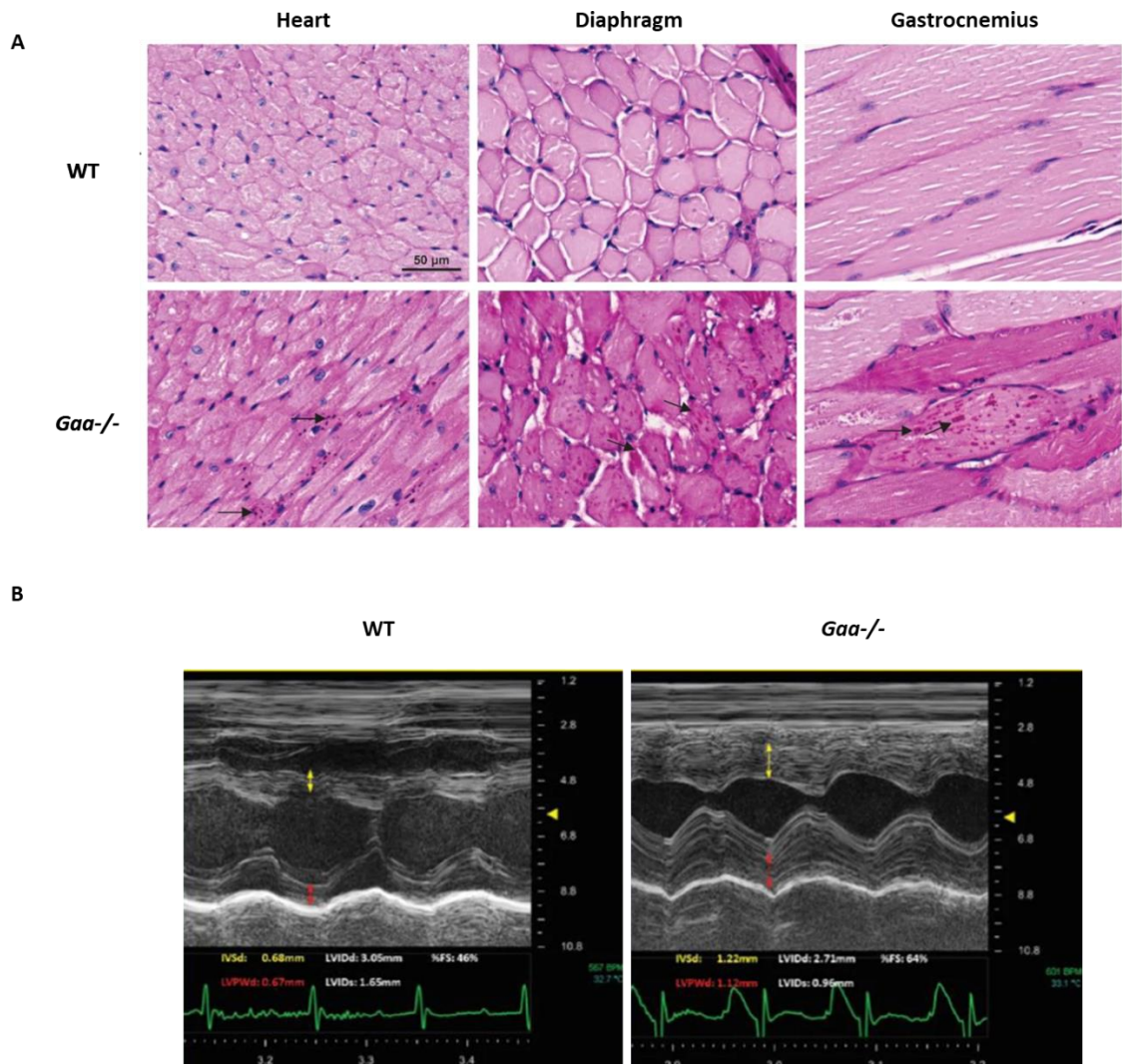
Figure adapted and modified from *Huang et al.*

*Figure 1.15 In vitro mouse model and hallmarks of PD. A) (PAS) staining of heart, diaphragm and gastrocnemius muscle from WT and Gaa-/-. Clear difference in glycogen accumulation, aggregates of PAS and severe irregularity in myofiber cells shapes in Gaa-/- vs WT. B) Echocardiography showing in 3moths old Gaa-/- mice substantial thickening of left ventricular posterior wall diameter (LVPWd) (red arrows) and interventricular septal diameter (IVSd) (yellow arrows) compared to coeval WT mice. Figure adapted and modified from Huang et al. (Huang et al., 2020)*

**1.6. Aims and objectives**

This study proposes to provide an endogenous stable supply of rhGAA through *ex vivo* lentiviral vector mediated HSC gene therapy and improve the cross correction in skeletal muscles in M6PR-independent manner by using MHPs driven entry. Lentiviral vectors will be used where appropriate as indicated in objective section.

We hypothesise that hGAA transgene under the transcriptional control of a mammalian promoter will be sufficiently expressed and followed by rhGAA protein production, secreted by transduced HSC progeny and uptaken by target cells i.e., skeletal, heart muscle and brain tissue to correct *in vivo* a mouse model of PD.

The objectives to accomplish the above aims are as follow:

1. Engineering hGAA transgene in order to improve enzyme secretion and skeletal muscle uptake. For this purpose, I tested a new in silico secretion peptide (SSP) (Guler-Gane et al., 2016) compared to the secretion peptide of insulin-growth factor II (IGFII) in use in our laboratory constructs and a muscle homing peptide (MHP) named as NewTag cloned as a single domain or in tandem, upstream of GAA. Various versions of hGAA transgene cassette were cloned into pCCL plasmid followed by lentiviral particles production and their usage *in vitro* to transduce appropriate cell lines.

2. Investigation of the functional capability of the test and control constructs in a stable gene edited GAA knockout C2C12 cell line, as an *in vitro* model of PD.

3. Assessment of the efficiency of uptake and the cellular localization of newly generated rhGAA proteins in both *Gaa-/-* C2C12 cells and H2K 2B4 cells in a transwell co-culture system.

4. Assessment of glycogen reduction in *Gaa-/-* C2C12 myofibers treated with different forms of rhGAA using *in vitro* transwell co-culture

5. Assessment of uptake and glycogen reduction in multiple different tissues and *in vivo* using a Pompe (*Gaa-/-)* mouse model with the use of lentiviral transduced *ex-vivo* HSC gene therapy.

6. Investigate MHP-receptor entry mechanisms and binding in muscle cells, by using culture inhibitory conditions applied to muscle cells and pathway analysis through total RNA sequencing.

# Chapter 2  Materials and methods

## 2.1.    Cloning strategies

### 2.1.1.  Design and synthesis of double stranded DNA fragment

Native sequence of human GAA (CCDS32760.1) was modified at N-terminus in order to improve secretion and uptake in skeletal muscle. Utilising muscle homing peptides RRQPPRSISSHP (Gao *et al.*, 2014), and a strong signal peptide MWWRLWWLLLLLLLLLWPMVWA (Guler-Gane et al., 2016) to test in the contest of hGAA. The amino acid sequences were reverse transcribed by using Bioinformatics (https://www.bioinformatics.org/sms2/rev_trans.html). Then, genetic constructs (Table 2.1) were designed using Serial Cloner 2.6.1 version (http://serialbasics.free.fr/Serial_Cloner.html) to contain the signal peptide, the muscle driving tag and an overlapping sequence to GAA for cloning strategy. Double stranded DNA fragments were then codon optimised for human expression and provided as dry pellet at 500nM concentration by Integrated DNA Technology (IDT) (https://eu.idtdna.com/codonopt ). Pellets were then resuspended in 10µl TE buffer (7.5 pH) to obtain a solution of 50nM/µl.

**Table 2.1. Designed constructs and PCR primer sequences.**

| Constructs ID | DNA sequence | Primers | |
| --- | --- | --- | --- |
| | | FW | RV |
| NewTag I-II GAA50 | TCCATCTCGTGCATGATCTACGTGCGTCAC ATGCAGTACAGCGCTGCCACCATGGGCAT CCCCATGGGCAAGAGCATGCTGGTGCTGC TGACCTTCCTGTGCTGTATCGCCGTGAGG AGGCAACCGCCTCGGAGCATTTCTAGCCA TCCAGCTAGCAGGAGACAGCCACCTCGGA GCATCAGCTCCCACCCAGAGACACACCCT GCACACCAGCAGGGAGCCTCCCGGCCCG GCCCTCGCGATGCACAGGCACACCCAGGC CGGCCTTTCAGTAGACCGCTGTTGAGCTG TCAGCACTACTAACTTGCGGTCAGT | TCCATCTCGTGCATGATCTA | ACTGACCGCAAGTTAGTAGT |
| NewTag I-II GAA55 | TCCATCTCGTGCATGATCTACGTGCGTCAC ATGCAGTACAGCGCTGCCACCATGGGCAT CCCCATGGGCAAGAGCATGCTGGTGCTGC TGACCTTCCTGTGCTGTATCGCCGTGAGG AGGCAACCGCCTCGGAGCATTTCTAGCCA TCCAGCTAGCAGGAGACAGCCACCTCGGA GCATCAGCTCCCACCCAGCACACCAGCAG GGAGCCTCCCGGCCCGGCCCTCGCGATGC ACAGGCACACCCAGGCCGGCCTTTCAGTA GACCGCTGTTGAGCTGTCAGCACTACTAA CTTGCGGTCAGT | TCCATCTCGTGCATGATCTA | ACTGACCGCAAGTTAGTAGT |
| NewTag I GAA70 | TCCATCTCGTGCATGATCTACGTGCGTCAC ATGCAGTACAGCGCTGCCACCATGGGCAT CCCCATGGGCAAGAGCATGCTGGTGCTGC TGACCTTCCTGTGCTGTATCGCCGTGAGG AGACAGCCACCTCGGAGCATCAGCTCCCA CCCAGCACACCCAGGCCGGCCTTTCAGTA GACCGCTGTTGAGCTGTCAGCACTACTAA CTTGCGGTCAGT | TCCATCTCGTGCATGATCTA | ACTGACCGCAAGTTAGTAGT |
| NewTag I 3 AA GAA70 | TCCATCTCGTGCATGATCTAAGCGCTGCCA CCATGGGCATCCCCATGGGCAAGAGCATG CTGGTGCTGCTGACCTTCCTGTGCTGTATC GCCGTGAGGAGACAGCCACCTCGGAGCA TCAGCTCCCACCCAGGAGCCCCAGCACAC CCAGGCCGGCCTTTCAGTAGACCGCTGTT GAGCTGTCAGCACTACTAACTTGCGGTCA GT | TCCATCTCGTGCATGATCTA | ACTGACCGCAAGTTAGTAGT |
| NewTag I 13 AA GAA70 | TCCATCTCGTGCATGATCTACGTGCGTCAC ATGCAGTACAGCGCTGCCACCATGGGCAT CCCCATGGGCAAGAGCATGCTGGTGCTGC TGACCTTCCTGTGCTGTATCGCCGTGAGG AGACAGCCACCTCGGAGCATCAGCTCCCA CCCAAATTTAAATGGCTTGGCTGCGGTTCT TAAGGGAGCCCCAGCACACCCAGGCCGG CCTTTCAGTAGACCGCTGTTGAGCTGTCA GCACTACTAACTTGCGGTCAGT | TCCATCTCGTGCATGATCTA | ACTGACCGCAAGTTAGTAGT |
| NewTag I-II GAA70 | TCCATCTCGTGCATGATCTACGTGCGTCAC ATGCAGTACAGCGCTGCCACCATGGGCAT CCCCATGGGCAAGAGCATGCTGGTGCTGC TGACCTTCCTGTGCTGTATCGCCGTGAGG AGGCAACCGCCTCGGAGCATTTCTAGCCA TCCAGCTAGCAGGAGACAGCCACCTCGGA GCATCAGCTCCCACCCAGCACACCCAGGC CGGCCTTTCAGTAGACCGCTGTTGAGCTG TCAGCACTACTAACTTGCGGTCAGT | TCCATCTCGTGCATGATCTA | ACTGACCGCAAGTTAGTAGT |
| NewTag I-II 3 AA GAA70 | TCCATCTCGTGCATGATCTACGTGCGTCAC ATGCAGTACAGCGCTGCCACCATGGGCAT CCCCATGGGCAAGAGCATGCTGGTGCTGC TGACCTTCCTGTGCTGTATCGCCGTGAGG AGGCAACCGCCTCGGAGCATTTCTAGCCA TCCAGCTAGCAGGAGACAGCCACCTCGGA GCATCAGCTCCCACCCAGGAGCCCCAGCA CACCCAGGCCGGCCTTTCAGTAGACCGCT GTTGAGCTGTCAGCACTACTAACTTGCGG TCAGT | TCCATCTCGTGCATGATCTA | ACTGACCGCAAGTTAGTAGT |

| | | | |
|---|---|---|---|
| NewTag I-II 13 AA GAA70 | TCCATCTCGTGCATGATCTACGTGCGTCAC ATGCAGTACAGCGCTGCCACCATGGGCAT CCCCATGGGCAAGAGCATGCTGGTGCTGC TGACCTTCCTGTGCTGTATCGCCGTGAGG AGGCAACCGCCTCGGAGCATTTCTAGCCA TCCAGCTAGCAGGAGACAGCCACCTCGGA GCATCAGCTCCCACCCAAATTTAAATGGCT TGGCTGCGGTTCTTAAGGGAGCCCCAGCA CACCCAGGCCGGCCTTTCAGTAGACCGCT GTTGAGCTGTCAGCACTACTAACTTGCGG TCAGT | TCCATCTCGTGCATGATCTA | ACTGACCGCAAGTTAGTAGT |
| SSP NewTag I-II GAA70 | TCCATCTCGTGCATGATCTACGTGCGTCAC ATGCAGTACAGCGCTGCCACCATGTGGTG GAGACTTTGGTGGTTGCTTTTGCTGCTTTT GCTGCTGTGGCCAATGGTATGGGCCGCTG CTAGGAGGCAACCGCCTCGGAGCATTTCT AGCCATCCAGCTAGCAGGAGACAGCCACC TCGGAGCATCAGCTCCCACCCAGCACACC CAGGCCGGCCTTTCAGTAGACCGCTGTTG AGCTGTCAGCACTACTAACTTGCGGTCAG T | TCCATCTCGTGCATGATCTA | ACTGACCGCAAGTTAGTAGT |

## 2.1.2. PCR amplification and restriction reaction

To amplify double stranded fragments, PCR reaction was employed using 1ng of DNA template, 10µM for each FW and Rev primer (See table 1), 10 µl Q5® Hot Start High-Fidelity 2X Master Mix (New England Bioscience, M0494S) and $H_2O$ to a final volume of 20µl. Thermocycling conditions for PCR reaction were an initial denaturation at 95°C for 30 seconds followed by 35 cycles of a denaturation step at 98°C for 5 seconds, annealing step at 63°C for 20 second, extension step at 72°C for 30 second, and then a final extension cycle at 72°C for 2 min. PCR product was then purified using Monarch® PCR & DNA Cleanup Kit (New England Bioscience, T1030S) and quantified using nanodrop ND-1000 NanoDrop Spectrophotometer (NanoDrop, E112352).

PCR amplicon containing desired insert and lentiviral vector containing the native *hGAA* sequence (pCCL.LCR-EFS.hGAA) were cleaved with two restriction enzymes: A*fel* (New England BioLabs, R0652S 10,000 units/ml) and F*sel* (New England BioLabs, R0588S 2,000 units/ml), to generate an insert of 250bp and a vector backbone of 13kb with compatible ends. Precisely, a total of 1µg of PCR product was added to 2µl CutSmart® Buffer (X10) (Neb, B7204S), 1ul of each restriction enzyme and $H_2O$ up to a final volume of 20µl. Similarly, 1µg of lentivirus vector was treated in 2µl SmartCut buffer(X10) (Neb), 1 ul of each restriction enzyme and $H_2O$ up to a final volume of 20µl. Both restriction reactions were kept at 37°C for 1 hour. Digested DNA bands were discriminated by electrophoresis

in 2% agarose gel and visualised on UV illuminator. The bands of interest were cut out from the gel and Monarch® DNA Gel Extraction Kit (New England Bioscience, T1020S) was used to purify DNA fragments from the agarose gel followed by Nanodrop quantification of the DNA.

## 2.1.3 Assembly of transfer plasmid and bacterial transformation

The ligation reaction of lentiviral backbone and desired insert was calculated by using NEBioCalculatorTM v1.10.1 ([https://nebiocalculator.neb.com/#!/ligation](https://nebiocalculator.neb.com/#!/ligation) ) to have 1:3 molar ratio of vector: insert. Generally, 50ng of lentiviral backbone and 2.3ng of insert were used in a ligation reaction. To those, 1µl of T4 DNA Ligase (New England Bioscience, M0202), 2µl T4 DNA Ligase Buffer (10X) and $H_2O$ up to a final volume of 20µl were added and gently mixed and left for overnight incubation at 16°C. Subsequent day, the ligation reaction was used for bacterial transformation.

One Shot™ Stbl3™ Chemically Competent E. coli (Thermofisher, C737303) were slowly thawed on ice for 5 minutes and promptly mixed with 5µl of ligation reaction. The transformation reaction was carried out on ice for 30 minutes prior to the heat shock at 42°C for 45 second. Then, transformation reaction was transferred on ice to rest for additional 3 minutes, before the addition of 250µl of S.O.C. medium. These were then incubated at 37°C for 1 hour with shaking at 250rpm. Once bacteria recovered, 100ul of transformation reaction were streaked onto pre-warmed LB agar plate containing 100µg/ml kanamycin and left overnight at 37°C. Subsequently, individual colonies were picked and placed in 5ml of LB broth with appropriate antibiotic for overnight incubation at 37°C, shaking at 250rpm. Plasmid DNA was extracted using Monarch® Plasmid Miniprep Kit (New England BioLabs, T1010L) and the concentration of DNA was evaluated by ND-1000 NanoDrop Spectrophotometer (NanoDrop, E112352). Restriction enzymes B*amHI*-HF® (New England BioLabs, R3136T 100,000 units/ml) and A*feI* (New England BioLabs, R0652S 10,000 units/ml) were employed to discriminate positive colonies for correct insert of interest. After 1 hour of digestion, samples were ran on 2% agarose gel and visualised using UV light. In parallel, undigested DNA samples were sent for Sanger sequencing to Source Bioscience England (Source Bioscience, Nottingham,

UK). The correct insert-lentiviral backbone clones were scaled up to 1L LB broth and processed for Maxi prep (QIAGEN, 12165) DNA purification, which were then used as transfer plasmids in the lentivirus production.

## 2.2 Lentivirus vector production, titration and VCN assessment.

### 2.2.1   Lentivirus production

Lentiviral vectors were obtained by transient transfection of $40 \times 10^6$/flask HEK-293T cells with second generation packaging plasmids and transfer plasmids containing the transgene cassette. Packaging plasmids consist of the envelop expressing plasmid, pMD2.G (Addgene), which carries a VSV-G codifying sequence, and GAG/POL plasmid, pCMVR8.74 (Addgene), containing HIV-1 gap, pol, tat and rev sequences which are essential in virion assembly. Packaging and lentiviral transfer plasmids were incubated together with $1 \times 10^{-7}$ mol/l of polyethylenimine (Sigma, 408727) for 30 minutes in Opti-MEM reduced serum medium (Gibco™, 11520386). HEK-293T cells were washed with 1x Phosphate Buffered Saline (PBS) (Gibco™, 14190144) to remove any residues of medium and serum before adding the transfection mix and left for 4h at 37°C in 5% $CO_2$ incubator. Then, transfection mix was replaced with 10% Fetal Bovine Serum (FBS) (Gibco™, 26140079), 1% Penicillin-Streptomycin (P/S) (Gibco™, Penicillin-Streptomycin 10,000 U/mL, 15140122) in Dulbecco's Modified Eagle's Medium (DMEM) (Gibco™, 11574486), and cells were allowed to recover at 37°C in 5% $CO_2$ incubator. The harvest of lentiviral particles was carried out 48 and 72 hrs post transfection. Collected medium containing viral particles was passed through 0.22µm pore cellulose filter stericups (MERC Millipore, S200B05RE) and concentrated by ultracentrifugation at 23,000 g for 2 hours. The viral particles were re-suspended in OPTI-MEM (Thermofisher Scientific, 31985062), left in ice for 30 minutes, pooled and stored at -80°C. The analysis of infectious units was done by titration of viral particles using HEX-293T cells follow by RT qPCR.

## 2.2.2 Lentivirus titration and assessment of vector copy number (VCN)

Lentiviral titration and assessment of VCN in transduced cell lines employed identical set of qPCR probes and primers. Separate subchapters depict description of cell seeding, culture and collection employed in aforementioned applications follow by common for both applications the qPCR component assembly, cycling process and finally primers and probes used to target genes of interest.

### 2.2.2.1 Lentiviral titration

24-well plate was seeded with HEX-293T cells at a density of $1x10^5$ cells/well and left for 24h to recover at $37^oC$ in 5% $CO_2$ incubator. The following day a preparation of 1:5 serial dilution of viral particle stock was made in a final volume of 200ul of OPTI-MEM. From each dilution 100µl of mixture was transferred to 400µl of 10% FBS and 1%P/S DMEM and added to a separate well with HEK-293T cell. Cells were then transferred into an incubator at $37^oC$ and 5% $CO_2$ and left for 72 hrs. After this period, cells were collected by trypsinization, individual well at a time, and genomic DNA extracted for qPCR.

### 2.2.2.2 VCN assessment in transduced cell lines.

Cell seeding, culture and collection depended on experimental settings/ expected goals and is described for each experiment/method separately in results and method sections respectively. In general, various cell lines or HSC were transduced with LV particles appropriate to *in vitro* or *ex vivo* outcomes. Where appropriate, transduced cell lines were harvested, washed with PBS, pelleted down and stored at $-80^oC$ or processed immediately for gDNA purification.

### 2.2.2.3 Assessment of lentiviral titration and vector copy number by qPCR.

Genomic DNA from transduced cells was extracted using DNeasy Blood and Tissue kit (Qiagen, 69504) in a final volume of 50µl. Each qPCR reaction consisted of 4µl of original

genomic DNA, 1µl mixture of primers and probe, conjugated to FAM reporter, for WPRE or GAA detection in the lentiviral sequence, 1µl mixture of primers and probe, conjugated to HEX reporter, for β-actin detection in the cell genome, 4µl of pure $H_2O$ and 10µl qPCR PrimeTime® Gene Expression Master Mix (IDT, 1055772) in a final volume of 20µl. Table 2.2. shows primers and probes sequence used for qPCR. Primers and probes sequences were customized and synthetized by IDT. The qPCR reaction ran for 1 cycle at 95°C for 5 min followed by 35 cycles of 65°C for 15 seconds and 72°C for 1 min in CFX96 Touch Real Time PCR Detection system (BioRad). The number of integrated copies of a lentiviral vector in the genomic DNA was extrapolated from a standard curve of known copy numbers of the same gene of interest. For quantify the vector copy number per cell, the mean quantity value of WPRE was divided by the mean value of β-actin in each sample.

**Table 2.2. Sequences of primers and probes used for qPCR.**

| Gene ID | Forward | Reverse | Probe |
|---|---|---|---|
| Viral WPRE | GCTTCCCGTATGGCTTTCA | GTTGCGTCAGCAAACACAG | /56-**FAM**/TTATGAGGA/ZEN/GTTGTGGCCCGTTGT/3IABkFQ/ |
| Human β-actin | GCTAAGTCCTGCCCTCATTT | GATGTCCACGTCACACTTCA | /5**HEX**/ AGTTTCGTG/ZEN/GATGCCACAGGACTC/3IABkFQ/ |
| Human GAA | TCTTTCCACCCTCTTACCCTAG | GTTCTCTGTCTCCATCATCACG | /56-**FAM**/CACGCACCA/Zen/CACCCACATTCTTTC/3IABkFQ/ |

## 2.3 Cell culture

### 2.3.1 K-562 (ATCC® CCL-243™)

Immortalised myelogenous leukemia cell line (K562) was cultured in Roswell Park Memorial Institute Medium (RPMI 1640) (Thermofisher, 12633012) containing 10% FBS (Thermofisher, 16140071) and 1% P/S (Thermofisher, 15070063). K562 cell line, due to its myelogenous origin, was chosen to work with as LCR enhancer, used in this study, exhibits its enhancing activity upon the promoter mainly in erythroid-like cells.

### 2.3.2 HEK-293T cells

Several human cells lines HeLa, HT1080, TE671 have been employed to produce lentivirus particles however these provided low viral titer (Macauley, 2016; Martinez-Molina et al., 2020; Merten et al., 2016; Tran & Kamen, 2022). Also monkey derived cells COS-1 were used showing high titer and reduced level of contaminating cellular proteins in supernatants (Pauwels et al., 2009). However, COS-1 cells are obligatory adherent cells and large-scale production is very difficult and costly. Furthermore, most of the current lentiviral vector production involves Human embryonic kidney 293T (HEK293T). choosing them over its progenitor HEK293 cell line as HEK 293T line stably express the SV40 large T antigen, which allow higher cellular proliferation rate, greater transfection stability and efficacy giving higher titre of produced viral particles comparing to previously mentioned human cell lines. In contrast to COS-1 cells the HEK293T cells can be adapted to grow in serum-free medium as suspension culture which is of paramount interest for large-scale vector productions (Merten et al., 2016).

HEK293T cells were cultivated in Dulbecco's modified Eagle's medium (DMEM) (Thermofisher, 41965039) containing 10% FBS and 1% P/S at 37°C in 5% $CO_2$ incubator.

### 2.3.3 H2K 2B4 cells

H2K 2B4 myoblast, immortalised mouse satellite cell-derived cell-line differentiated *in vitro* to form myotubes, were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 20% FBS, 2% chicken embryo extract (Sera laboratories international, CEE), 4 mM L-glutamine (Gibco™,L-Glutamine (200 mM), 25030081), 20 U/ml of γ-IFN (Chemicon, IF-002  and 1% P/S at 33°C in 5% $CO_2$. Upon differentiation into myofibers H2K 2B4 were incubated in DMEM supplemented with 5% Horse Serum (Thermofisher, 16050122), 4 mM L-glutamine, 1% P/S and transferred at 37°C in 5% $CO_2$ incubator.

### 2.3.4 C2C12 (ATCC® CRL-1772™)

Mouse musculus myoblast C2C12 cells were cultured in DMEM with 10% FBS and 1% P/S at a density not exceeding 25% confluence in a given culture flask. To stimulate cells differentiation into myofibers C2C12 were seeded at high density ($2\text{x}10^5$ per well in 24 well plate) and cultured in DMEM supplemented with 5% horse serum and 1% P/S. C2C12 wild type myoblasts served as a platform to generate stable *Gaa -/-* myoblast line using CRISPR/Cas9 system.

### 2.3.5 Cell culture general conditions

All cell cultures were maintained at 37°C in 5% $CO_2$ incubators and subcultures were performed using 0.25% Trypsin-EDTA for adherent cell lines, unless otherwise stated.

### 2.4 Enzymatic activity assessment techniques

### 2.4.1 Acid alpha-glucosidase (GAA) enzymatic activity assay

The 4MU-G is used as a fluorogenic substrate for GAA enzyme. GAA enzymes present in tested sample/s cleaves off the 4MU part of substrate while converting alpha-D-glucoside into glucose (Figure 2.1). Released 4MU emits light which than can be measured and quantified. Higher fluorescent signal indicates greater content of GAA enzymes in the sample. This method allows for the identification, characterization, and kinetic analysis of GAA enzyme.

Control or transduced K562 cells were harvested, washed with PBS, pelleted down and lysed using lysing buffer containing: 50mM Tris/HCL, pH7.4, 150mM NaCl, 5mM EDTA, 1% cocktail protease inhibitor and 1% Triton X-100. Samples were transferred in ice for 10 min and span at 3000rpm for 5 min to remove any cell debris from supernatant. 200mM Acarbose stock (Sigma, A8980-1G) was diluted in $H_20$ to obtain a working concentration of 7.5μM. Subsequently, 5μl cell lysate from each sample and 10μl of

7.5µM Acarbose were transferred in duplicates into 96-well black plate flat bottom, followed by the addition of 10µl 4mM 4-Methylumbelliferyl-α-D-glucopyranoside (4MU-D) (MELFORD, M65700-1.0) substrate diluted in McIlvane buffer, pH3.8 and incubation at 37°C with gentle shaking for 1 hour. The enzymatic reaction was terminated by addition of 175µl 0.25M glycine stopping buffer, pH 10.4. Fluorescence signal was detected at 360excitation/460emision filters using OPTIMA plate reader. Fluorescence readouts of samples with unknown GAA activity were extrapolated from a standard curve of know concentration of 4MU. The GAA activity was expressed as nmol of 4MU/h normalized by mg of proteins in the lysate, determined by Bradford assay.
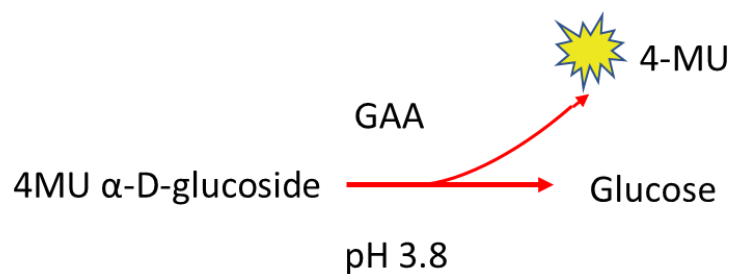


**Figure 2.1. GAA enzymatic activity.** *Breakdown of 4MU-D fluorogenic substrate producing glucose and fluorescent dye 4MU.*

### 2.4.2 Glycogen assay

To investigate glycogen concentration in treated C2C12 myofibers EZScreenTM Glycogen Colorimetric Assay Kit (384-well) was used (BioVision, # K960-400). Cell pellets were resuspended in 100 µl of pure $H_2O$, boiled for 10 min at 95°C and span at 6000 RPM for 5 min to remove debris. Supernatants were incubated with mix of enzymes provided in EZScreenTM Glycogen Colorimetric Assay Kit (384-well). After 2 hrs of incubation at 37°C enzymes hydrolysed the glycogen in each biological sample releasing glucose which is then specifically oxidized to generate a product that reacts with OxiRed Probe provided in the enzymes mix. The OxiRed probe was measured by colorimetric

assessment using OPTIMA Plate reader at 590nm wavelength. Values of unknown glycogen concentration in each sample were extrapolated from a standard curve of known concentration of glycogen. The glycogen content was expressed as µg/mg of total protein in the sample.

### 2.4.3 β- galactosidase (β-Gal) enzymatic activity assay

β-Gal is a lysosomal enzyme used in variety of settings to determine viability of cells preparation. Its synthetic substrate 4-methylumbelliferyl-**β**-D galactopyranoside also call 4MU is a fluorogenic substrate. Similarly, to 4-Methylumbelliferyl- **α**-D-glucopyranoside used in detection of GAA, 4-methylumbelliferyl-**β**-D galactopyranoside undergoes a cleavage in presence of β-Gal and emits light, which than can be measured and quantified.

K562 and 293T cells were harvested and processed as previously described in section 2.4.2, subsequently 10µl of 7.5µM Acarbose and 5ul of cell lysate, in duplicates for each cell fraction, were placed in parallel in two black 96 well plate. One set of duplicates were tested for GAA activity with 4-Methylumbelliferyl-**α**-D-glucopyranoside, the other set of duplicates were tested for β-Gal activity by adding 4-methylumbelliferyl-**β**-D galactopyranoside (Sigma Aldrich, M -1633). Each reaction was provided with 10µl of substrates to reach 4mM concentration in the final reaction volume. Samples were left at 37°C in gentle shaking for 1 hr. Termination of enzymatic reaction, fluorescent reading and calculation of enzymatic activity using known 4MU standards was processed for both enzymes as previously described in section 2.4.2

### 2.5 Concentration and purification techniques

### 2.5.1 Concentration of protein in media of K562 cells

Medium from control or transduced K562 cells, bearing divergent variants of GAAmCherry transgene, were harvested and concentrated using Pierce™ Protein

Concentrators PES, 3K MWCO, 5-25ml (Thermofisher, 88526). Concentration of mediumallows to remove access of water and any unwanted molecules which are smaller than the protein concentrator's threshold, in this case 3kDa, allowing to concentrate GAA enzyme. 50ml of cell medium was concentrated to a volume of 1ml and stored at -80°C.

### 2.5.2 Harvest of glycoproteins using HiTrapTM Con A 4B

A total of 50 ml of cell medium was harvested from control or transduced K562 cells after 5 days of culture and glycoproteins from medium were captured by employing HiTrap$^{TM}$ Con A 4B 1 ml Column (GE Healthcare Life Science, 28952085). Prior to run medium, the columns were activated by applying 5x column volumes (CV, 1ml) of binding buffer containing 20mMTris-CHL, 0.5M NaCl, 1mM $CaCl_2$, pH7.4. Medium was pumped onto Con A column using a peristaltic pump at a low flow rate, 0.5ml/min. Subsequently, columns were washed with 10x CV binding buffer, followed by dilution of glycoproteins using 5x CV of elution buffer, consisting of 0.3 M methyl-α-D-glucopyranoside, 20 mM Tris-CHL, 0.5 M NaCl, pH7.4 were prepared, at 0.5ml/min flow rate. Samples were stored at -80°C for further analysis.

### 2.5.3 Purification of Histidine-tagged proteins using HisTrap HP

293T cells grown in T175 culture flasks were transiently transfected using 1ul transfecting reagent polyethylenimine (PEI) (Sigma-Aldrich, 408727-100ML) mixed with 50ug of LV transfer plasmid containing SSP-NewTagI-II-Myc-His fragment in 10ml plain DMEM. Transfection mix was left on 293T cells for 3 hrs, then removed and 15ml DMEM supplemented with 5% FBS and 1% P/S was replenished. Cells were left for 3 days to recover and then, NewTagI-II-Myc-His protein was purified from medium using HisTrap$^{TM}$ affinity chromatography column (GE Heathcare 17-5247-01, HisTrap HP 1ml). Prior to His-tagged protein purification, HisTrap column was washed with 5 column volumes of ddH2O followed by activation with binding buffer containing 20mM sodium phosphate (Sigma-Adrich, 342483), 0.5M NaCl (Sigma-Adrich, S9888-500G) and 30mM

imidazole (Sigma-Adrich, I2399-100G) with final pH at 7.4; 5 column volumes of binding buffer were applied. Medium containing His-tagged protein were run through HisTrap column at 1ml/min. Subsequently, 5 column volumes of binding buffer were applied to HisTRap column to remove any residual medium. Finally, 5 column volumes of elution buffer, containing 20mM sodium phosphate, 0.5M NaCl and 500mM imidazole at pH 7.4 was applied to elute NewTagI-II-Myc-His protein. Eluted protein was immediately stored at - 80°C. NewTagI-II-Myc-His tag protein was used in downstream assessment of NewTag cellular entry point/mechanism.

## 2.6 Confocal microscopy

### 2.6.1 Confocal imaging of rhGAA

$1 \times 10^5$ H2K 2B4 cells were seeded in 8 well glass bottom cell culture chambers, 15µ-Slide (Ibidi, 80827), and let differentiate in myotubes at 37°C in 5% $CO_2$ incubator. After 3 days, H2K 2B4 multinucleated myofibers were treated for 24 hours with 1uM rhGAA, cloned in frame with mCherry and concentrated from transduced K562 medium. The following day, LysoTracer® (Thermofisher, L12492) for live cells was added to a final concentration of 50nM per well 1h prior to imaging. Cells were then washed twice with PBS and 200µl 10%FBS and 1% P/S DMEM was replenished per well. Samples were processed for live cell imaging at Inverted Zeiss LSM 710 confocal microscope. LysoTracer® was excited by 488 nm and mCherry by 561 nm wavelength laser. In addition, a heat chamber at 37°C was used to keep H2K 2B4 myofibers in optimal conditions while imaging.

### 2.6.2 Confocal imaging of CI-M6PR

Both wild type and *Gaa-/-* C2C12 myoblast were differentiated into myofibers by seeding $2 \times 10^5$ per well into 4 well glass bottom cell culture chambers, 15µ-Slide (Ibidi, 80426) at 37°C in 5% $CO_2$ incubator for 4 days. Upon differentiation into myofibers, cells were fixed with 4% PFA for 10 min, washed with PBST (containing 0.05% Tween-20),

blocked with 5% rabbit serum for 1h and incubated with rabbit anti-CI-M6PR at 1:1000 dilution at RT for 1 hour. Cells were then washed 4 times for 5 min each in 1xPBST and incubated for 1 hour at room temperature with Goat anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 488 (Thermofisher, A-11008) at 1:1000 dilution. Samples were then washed other 4 times for 5 min each in PBST and treated with Alexa Fluor™ Plus 405 Phalloidin (Thermoffisher, A30104) at a final concentration of 50nM to stain actin filaments of C2C12 myofibers. SYTOX™ Green Nucleic Acid Stain (Thermofisher, S7020) for nucleic acid staining at a recommended concentration of 5mM in final volume of 500µl PBS was added prior to confocal imaging. Each well was then washed 2 times for 5 min each with 1xPBST to remove any dye in excess. Samples were processed for cell imaging at Inverted Zeiss LSM 710 confocal microscope. Phallodin 405 was excited by 405 nm, SYTOX™ Green by 488 nm and Alexa 647 by 633 nm wavelength laser.

## 2.7 Western Blotting

### 2.7.1 Western Blotting SDS-PAGE immunoblotting

Total protein from cell pellets were extracted using cell lysing buffer containing: 50mM Tris/HCL, pH7.4, 150mM NaCl, 5mM EDTA, 1% cocktail protease inhibitor (Sigma-Aldrich, P8340-5ML) and 1% Triton X-100. Cell pellets were individually resuspended in lysing buffer and left in ice for 5 min follow by centrifugation at 13,000g to remove cell debris. To analyse the total amount of protein concentration in each lysate Pierce™ BCA Protein Assay Kit (ThermoFisher, 23225) was used. Values of absorbance from unknown concentration from each sample were plotted against a standard of known protein concentration (BioRad, 5000206) using GraphPad Prism software (GraphPad Prism 8.3.0). Exact volume of each sample containing 20µg of proteins was calculated, mixed with NuPAGE™ LDS Sample Buffer x4 (Thermofisher, NP0008) and H$_2$O up to a final volume of 25µl. Samples were boiled for 5 min at 90$^\circ$C. Subsequently, samples were loaded on NuPAGE™ 4-12% Bis-Tris sodium dodecyl sulphate polyacrylamide separating SDS-PAGE Protein Gels (Thermofisher, NP0321BOX). After electrophoresis, proteins

were transferred to polyvinylidene fluoride membrane (PVDF) (Bio-Rad Laboratories, 1620177), blocked for 1hour with 5% milk in 1xPBST containing 0.05 % Tween-20 or 5% BSA (Sigma-Aldrich, A4503-100G ) in  1xTBS (Tris-buffered saline) (Sigma-Aldrich, 524753-1EA) supplemented with 0.05% Tween-20 for phosphoproteins and incubated with rabbit monoclonal anti human GAA antibody (Abcam, ab137068), rabbit monoclonal anti mouse CI-M6PR (Abcam, ab124767), rabbit monoclonal anti AP2 (Abcam, ab76007) or rabbit monoclonal anti AP2-P (Abcam, ab109397) at 1:1000 dilution and left at 4$^o$C overnight. Subsequently, membranes were washed 4 times, 5 min each, in 1xPBST or 1xTBS and then incubation with Goat anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, horseradish peroxidase (HRP) (Invitogen, G-2123) conjugated, at 1:1000 dilution in 5% milk PBST or 5% BSA TBS at RT for 1 hour. Enhanced chemiluminescence ECL SuperSignal™ West Pico PLUS Chemiluminescent Substrate (Thermofisher, 34579) was added to the membranes and Bio-Rad ChemiDoc XRS system was used to detect and image chemiluminescent signals.

### 2.7.2 ImageJ analysis of Western blotting

The intensity of Western blot bands in developed membranes was analysed and quantified using ImageJ software (ImageJ 1.51j8 Wayne Rasband, National Institute of Health, USA). Areas of interest were marked, and the software calculated a total band's intensity for each individual band of interest. Bands signals were acquired by measuring an area calculated in 360x1768 pixels, 8-bit picture and then normalised by VCN of the individual samples.

### 2.8 Software

Graphpad Prism v8.3.0: Arrangement of data into appropriately formatted graphs, with subsequent statistical analysis (https://www.graphpad.com/scientific-software/prism/).

Serial Cloner v2.6.1: Design gene map and restriction digest design analysis. Furthermore, this software was used to analyse Singer sequence and gene alignments (http://serialbasics.free.fr/Serial_Cloner.html ).

ImageJ v1.51j8: This software was used to analyse picture data from confocal microscopy and quantification of band intensity on WB in this study (https://imagej.nih.gov/ij/download.html ).

FlowJo v10.8: Analysis of FCS files used for all flow cytometry (https://www.flowjo.com/solutions/flowjo/downloads ).

## 2.9 Analyses carried on transduced K562 cells

### 2.9.1 Assessment of post-translational modification in rhGAA protein secreted into cell medium

Glycoproteins purified by HiTrap ConA 4B from medium of K562 cells bearing different rhGAA constructs were used to evaluate glycosylation patterns on rhGAA proteins. About 20µg of glycoproteins were diluted in 1 µl of Glycoprotein Denaturing Buffer (10X) and $H_2O$ up to 10 µl total volume and then denaturised at 100°C for 10min and transferred on ice for 1 min. Subsequently, 2µl of 10x GlycoBuffer 3, 3µl of Endo H (New England Bioscience Neb P0702S 500,000 units/ml) or 1µl of PNGase F (New England Bioscience Neb P0704S 500,000 units/ml) enzyme and $H_2O$ up to a final volume of 20µl were gently mixed with 10µl of denatured samples. Enzymatic reactions were incubated at 37°C for 1 hour, stopped by heat denaturation at 90°C for 5 min and stored at -80°C.

### 2.9.2 mRNA level of acid alpha-glucosidase in transduced K562

$5\times10^6$ K562 cells were collected, and their total mRNA was extracted by using mRNA kit (Sigma, 11741985001). 20µl of total amount of mRNA from each sample was then converted to cDNA using SuperScript ® III First-Strand Synthesis System for RT-PCR (Invitrogen, 18080-051). Finally, 4µl of cDNA was combined with probes and primers for

rhGAA (Table 2.2) and the housekeeping gene β-actin. After running a qPCR following the protocol detailed in section 2.4.1., CT (cycle threshold) values of all samples were obtained and data were analysed by using ΔΔCT values equation, where the first ΔCT was acquired by subtracting CT value of β-actin gene from CT value of GAA gene, this operation was done for all samples including untransduced sample used as reference control. The second ΔCT was calculated by subtracting the ΔCT of the reference control from the ΔCT of each transduced sample as shown by this equation: ΔΔCT = ΔCT (transduced sample) − ΔCT (reference sample). Finally, the value from ΔΔCT was put to the power of 2 in **$2^{\wedge-(\Delta\Delta CT)}$**. Final step of gene expression assessment was a normalisation of **$2^{\wedge-(\Delta\Delta CT)}$** values by VCN from each individual sample.

## 2.10 C2C12 *Gaa-/-* cell line genesis and downstream analysis.

### 2.10.1 Generation of G*aa* knockout C2C12 cells using CRISPR Cas9 platform

Single guide RNAs (sgRNAs) used for the generation of GAA knockout C2C12 line were designed utilising the online Benchling CRISPR tools (https://benchling.com/crispr). Three guide RNAs targeting within exon 2 were chosen for maximal on-target and minimal off-target binding affinity scores (Table 2.3). The synthesis of chemically modified guides with 2'-O-methyl 3' phosphorothioate modification in the first and last nucleotides was performed by Synthego (SYNTHEGO, California, USA, www.synthego.com ).

**Table 2.3. single guide RNA (sgRNA) fragments.**

| Guide No | Sequence | Strand | On-Target Score | Off-Target Score |
|---|---|---|---|---|
| GAA sgRNA1 | CCTTATCTCTCTGACTACAG | sense | 72.6 | 40.6 |
| GAA sgRNA2 | CTGGGAGGGAAGAAACACCA | antisense | 71.5 | 46.6 |
| GAA sgRNA3 | CTGGCTTGTAACCTTCCTGG | antisense | 71.1 | 50.6 |

An active Cas9 protein (Alt-R® S.p. Cas9 Nuclease V3, 1081058) was bought from IDT. The assembly of sgRNA and Cas9 protein complexes were performed by mixing 1.2:1 molar ratio gRNA:Cas9 in a total volume of 10µl and left to rest at 37°C for 10 min. These complexes were then individually electroporated into C2C12 cells using NEON™ kit (Invitrogen, MPK1025), with specific electroporation parameters for this type of cells provided by NEON™ manufacturer (https://www.thermofisher.com/content/dam/LifeTech/migration/en/filelibrary/cell-culture/neon-protocols.par.42869.file.dat/c2c12-muscle.pdf) (Table 2.4). After electroporation, cells were plated in T175 flask in 10% FBS in DMEM medium without P/S and transferred at 37°C in 5% $CO_2$ incubator to allow their recovery.

**Table 2.4. Electroporation parameters for C2C12 cells.**

| Pulse voltage (v) | Pulse width (ms) | Pulse number | Cell desnsity (cells/ml) | Transfection efficiency | Viability | Tip type |
|---|---|---|---|---|---|---|
| 1,650 | 10 | 3 | $5\times10^6$ | 95% | 96% | 10µl |

Following the recovery, 1/4th of T175 flask from each treatment was harvest and DNA was extracted using DNeasy Blood & Tissue Kit (Qiagen, 69504) and sent for Sanger sequencing using Source Bioscience service (Source Bioscience, Nottingham, UK). Employing ICE (Inference of CRISPR Edits) platform (https://ice.synthego.com/#/ ), *Gaa* gene in wild type C2C12 cells was compared to the *Gaa* gene sequenced from the treated cells to determine the level of gene disruption caused by each gRNA-Cas9 complex. Two sgRNAs gave 85% and 95% mutation rate while the third sgRNA did not knock down *Gaa* gene. Then, C2C12 cells treated either with sgRNA1 and sgRNA2 were single-cell sorted in 96well plate to generate a single cell clonal expansion. After a week, 10 wells from each sgRNA treated C2C12 were individually transferred into T25 flask and allowed to furtherly grow and then expand into T75 flasks with no more than 30% confluence at any passage. Eventually, genomic DNA from all 20 T75 flasks, containing

clones of sorted C2C12 cells, were harvested for Sanger sequencing analysis. DNA sequences from each clone were compared to wild type C2C12 to determine acquired mutations using BLAST® platform.

## 2.10.2 Purification of membrane proteins for CI-M6PR assessment in WT vs. *Gaa* KO C2C12

Both wild type and *Gaa-/-* C2C12 myoblast were seeded in 6 well plate, at $2x10^6$ cells/well in two wells each, in 5% Horse Serum in DMEM and 1% P/S and left for 4 days to differentiate into myofibers. Cells were then collected using 0.25% Trypsin-EDTA, washed with PBS and span down at 300g for 5 min. Plasma membrane proteins were harvested using Mem-PER™ Plus membrane Protein Extraction Kit (Thermo Science, 89842). Cell pellets were resuspended in 0.75ml of Permeabilization buffer, vortexed, incubated on ice for 10 min and centrifuged at 16,000g for 15 min. Supernatant was carefully removed and pellets containing plasma membranes were resuspended in 0.5ml of Solubilisation Buffer, placed on ice for 30 min, followed by centrifugation at 16,000g for 15 min. Supernatants containing membrane proteins were collected and transferred at -80°C for further analysis.

## 2.10.3 Transwell co-culture: assessment of rhGAA uptake

A total of $2x10^5$ C2C12 *Gaa-/-* cells were seeded per well in 24 well plate coated with 0.1mg/ml matrigel (Thermofisher, A1413201) and allowed to grow/differentiate in 1ml of 10% FBS DMEM medium with 1% P/S for 4 days at 37°C in 5% $CO_2$ incubator. Upon differentiation into myofibers, medium was changed for 5% horse serum (Thermofisher, 16050122) in DMEM and 1% P/S, and 6.5 mm transwell with 0.4 µm pore polyester membrane inserts (Sigma-Aldrich, Corning® Transwell®, CLS3470) were added to each well. K562 cells were then seeded in 350µl DMEM containing of 5% horse serum and 1% P/S in the inserts. C2C12 *Gaa-/-* cells were exposed to K562 cell lines bearing different rhGAA constructs. Polyester membrane with 0.4µm and $1x10^6$ pores per membrane allows the interchange of only solutes present in cell medium, thereby only secreted

proteins from K562 cells including rhGAA enzymes were able to migrate through membranes and engage with myofibers. As detailed per each experiment, 5mM of D-Mannose 6-phosphate disodium salt hydrate (Sigma-Aldrich, M6876) used as a mannose 6-phospate receptor (M6PR) competitive antagonist was added to cell medium. Cells were maintained in co-culture from 24 to 72 hrs. Subsequently, inserts holding K562 cells were carefully removed, myofibers washed twice with PBS and harvested using 0.25% Trypsin-EDTA, span down at 300g and pellet washed once with PBS. Cells were span again at 300g and stored as dried pellets at -80°C.

### 2.10.4 Lysosomal glycogen level determination in starvation assay – *Gaa-/-* C2C12 cells

$2x10^5$ *Gaa-/-* C2C12 cells were seeded in transwell in 1ml of 10% FBS 1% P/S in DMEM medium at 37°C in 5% $CO_2$ incubator. Upon differentiation into myofibers, cells were washed once with PBS and 10% dialysed FBS (Thermofisher, A3382001), 1% P/S and glucose and pyruvate free DMEM (Thermofisher, 11966025) medium was replenished. Cells kept in medium free from glucose and pyruvate and with dialysed FBS were prevented from taking up any molecules which could be used as a substance to form glycogen or glucose and generate a background noise in a glycogen assay. The treated C2C12 cells were kept in starvation conditions for 24 hrs to use up all intracellularly available glucose. Consequently, medium was replenished with 350µl of 10% dialysed FBS, 1% P/S, glucose and pyruvate free DMEM with addition of 5mM M6P. K562 cell line expressing different GAA construct were counted and $1x10^6$ was placed in 350µl of 10% dialysed FBS, 1% P/S, glucose and pyruvate free medium inside transwell inserts. Co-culture was left for 48 hrs. Consequently, transwell was disassembled and *Gaa-/-* C2C12 cells were washed twice with PBS, detached and span down at 300g for 5 min. Supernatant was removed and dry cell pellets were frozen down and stored at -80°C for future glycogen assay analysis.

## 2.10.5 Inhibition of endocytosis

$2 \times 10^5$ Gaa-/- C2C12 cells were seeded in 10% FBS, 1% P/S in DMEM on matrigel coated 12 well plate and allowed to differentiate. Upon differentiation, Gaa-/- C2C12 myofibers were cultured in DMEM with 5% horse serum and 1% P/S. Series of pharmacological inhibitors, inhibitory conditions were applied individually per each well containing differentiated Gaa-/- C2C12. M6P inhibitor was also applied to each of the wells to block GAA uptake through M6PR, with exception for untreated C2C12 Gaa-/-, treated with only concentrated medium from untransduced K562 and C2C12 Gaa-/- treated with NewTagI-IIGAA serving as a positive control for GAA uptake. The pharmacological inhibitors and their working concentrations were as follow: 7.7µM sodium azide (Sigma-Aldrich, 71290-10G), 5µg/ml chloroquine (MP Biomedicals LCC, 193918), 2.5µM chlorpromazine (Sigma-Aldrich, C8138-5G), 75µg/ml fucoidan (Cayman Chemical, 20357), 2µg/ml dextran sulfate (Sigma-Aldrich, 75027-5G), 200µM genistein (Cayman Chemical, 10005167). In addition, temperature i.e., 4°C were used as an inhibitory condition for active transport. Cells were exposed to inhibitors 1 hour prior to addition of 1µM NewTagI-II GAA. Also, cells exposed to 4°C inhibition conditions were cultured for 1 hr in the above temperature prior to addition of aforementioned constructs. After addition of 1µM NewTagI-II GAA, cells were incubated at 37°C and 5%$CO_2$ for 24hrs; while cells exposed to 4°C condition were carefully wrapped in parafilm to preserve $CO_2$ and correct pH. After 24hrs incubation, samples were washed twice with PBS, harvested, pelleted down and transferred to -80°C. Cell pellets were then lysed and GAA activity measured to determine level of inhibition of GAA uptake.

## 2.10.6 Targeted inhibition of selected endocytic receptors

To establish NewTagI-II potential plasma membrane receptor, a selective inhibition of endocytic receptors identified through RNA-seq was undertaken. $1.9 \times 10^5$ Gaa-/- C2C12 cells were seeded per well in 24 well plate in 1 ml differentiation medium and placed at 37°C and 5% $CO_2$ until full differentiation into myofibers. Upon differentiation, 4 selective inhibitors: 1µM BLT-1 (Sigma-Adrich, SML0059-5MG), 1µM RAP (Sigma-Aldrich, 553506-M), 5µg/ml AMD3100 (Sigma-Adrich, A5602-5MG) and 3µM STI571

(Sigma-Aldrich, CDS022173-25MG) were added in separated wells to specifically inhibit, respectively. Cells were exposed to the inhibitor for 1-hour prior to stimulation with NewTagI-II tag protein. After this period, 80μl of NewTagI-II-Myc-His eluted protein was added to cell medium, gently mixed and left for 5 minutes. Alongside, 1 well with differentiated *Gaa-/-* C2C12KO, treated with NewTagI-II-Myc-His protein only was established as Treated Control. Furthermore, 1 well with untreated neither with NewTag protein nor inhibitor serving as Untreated Control. In the future, additional control elution buffer might be set up to determine potential impact, if any, of this buffer on C2C12KO cells. After 5 minutes all conditions including treated and non-treated controls were harvested, washed twice with ice-cold PBS, span at 300g for 5 minutes and cell pellets placed on dry ice. Samples were then processed for AP2-P protein level detection.

**2.10.7 RNA-seq analysis**

$2x10^6$ *Gaa-/-* C2C12 cells were seeded in 10% FBS, 1% P/S in DMEM on matrigel coated 6 well plate and allowed to differentiate. Upon differentiation, *Gaa-/-* C2C12 myofibers were cultured in DMEM with 5% horse serum and 1% P/S. 240μl elution containing NewTagI-II Myc His tag was added to differentiated myofirbes and left for 6 hrs in 37°C and 5% $CO_2$ incubator. After this time cells were washed twice with 10 ml PBS and harvested, and their total mRNA was extracted by using mRNA kit (Sigma, 11741985001). 100μl of eluted mRNA was given to UCL genomics to process. The entire RNA-seq process including library assembly, sequencing, end point gene alignments and gene set enrichment raw data processing was conducted by our collaborators in UCL genomics.

The overview of the process was as follow. Generation of cDNA from mRNA using KAPA mRNA HyperPrep Kit (Roche, KK8580). mRNA was isolated from total RNA by use of paramagnetic Oligo dT beads (Invitrogen, 61005) to pull down poly-adenylated transcripts. The purified mRNA was fragmented using chemical hydrolysis (heat and divalent metal cation). Strand-specific first strand cDNA was generated using Reverse Transcriptase. The second cDNA strand was synthesised using dUTP in place of dTTP, to

mark the second strand. R5 primers (Swift BioSciences p/n 66096) were used to amplify the library. Samples were denatured and sequenced on the NextSeq 500 instrument (Illumina, 20024913) using a 75bp single read run with a corresponding 8bp single sample index and 8bp unique molecular index read. Finally, run data was demultiplexed and converted to fastq files (text-based file containing biological sequence and its quality score) using Illumina's bcl2fastq Conversion Software v2.20 (https://emea.support.illumina.com/downloads/bcl2fastq-conversion-software-v2-20.html). Fastq files were then tagged with the unique molecular identifiers (UMI) read using UMITools (https://github.com/CGATOxford/UMI-tools ) and aligned to the mouse genome UCSC mm10 using RNA-STAR 2.5.2b (https://emea.illumina.com/products/by-type/informatics-products/basespace-sequence-hub/apps/rna-seq-alignment.html). Aligned reads were then UMI deduplicated using Je-suite software v1.2.1 to and reads per transcript were counted using FeatureCounts v2.0.3. (http://subread.sourceforge.net/) to produce a digital output of gene expression. Differential expression was estimated using the BioConductor package SARTools (https://rdrr.io/github/PF2-pasteur-fr/SARTools/), a DESeq2 (https://bioconductor.org/packages/release/bioc/html/DESeq2.html) and EdgeR wrapper (https://bioconductor.org/packages/release/bioc/html/edgeR.html).

## 2.11. Mice strain *in vivo* study

Both, genetically modified *Gaa* knockout mice B6;129-Gaatm1Rabn/J and wild type C57BL/6 were provided by The Jackson Laboratory (https://www.jax.org/ ).

## 2.12 *In vivo* assessment techniques

In vivo study of gene therapy for PD was conducted using C57BL/6 *Gaa-/-* mice as a recipient of gene therapy and wild type non genetically modify C57BL/6 mice.

4 experimental arms were set up to explore correction of gene therapy included on Table 2.5

**Table 2.5. *In vivo* experimental arms**

| Experimental arms | Number of mice per group | Genotype of animals |
|---|---|---|
| Wild type animals as a positive control | 8 | C57BL/6 |
| GAA-/- untreated mice as a negative control | 6 | C57BL/6 *GAA-/-* |
| pCCL SSP-NewTagI-IIGAA | 5 | C57BL/6 *GAA-/-* |
| pCCL GAA Native | 5* | C57BL/6 *GAA-/-* |

*due to death of 1 animal during this trial, only 4 participant remained.

The Experimental design and timeline are presented on Figure 2.2. End point functional test and tissues collections were conducted 3 months post-transplantation.
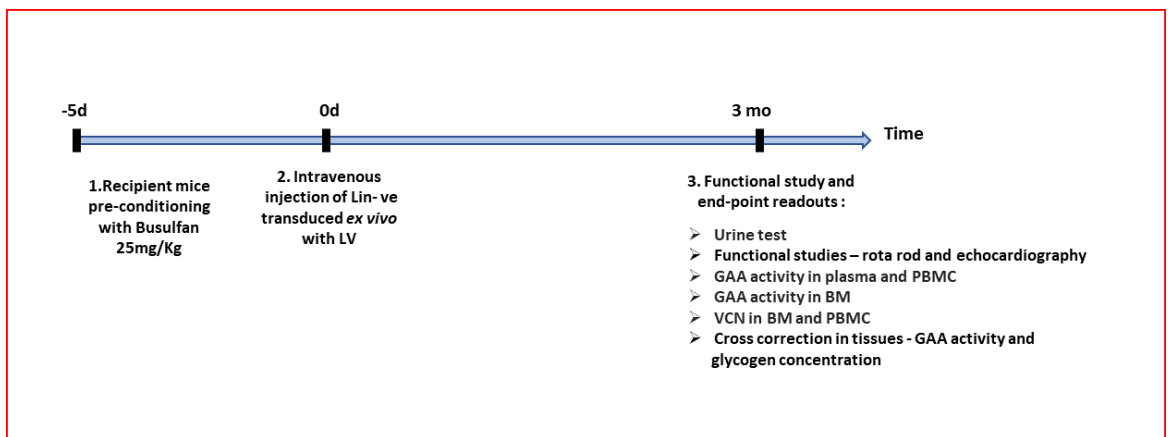


**Figure 2.2. Experimental timeline. 1***) Recipient female mice were preconditioned for 4 days with 25mg/Kg Busulfan in order to achieve complete myeloablation. **2) Lin-ve** harvested from male donors and ex vivo transduced with appropriate LVs were intravenously infused in preconditioned female recipients. **3)** 3 months post-transplant cross-correction study including functional correction and end point readouts.*

To determine VCN of infused Lin-ve into recipient mice a small sample of transduced Lin-ve was collected and split in two conditions *in vitro* cell culture:

1) Liquid culture - cultured for 7 days
2) Colony forming Units - cultured for 14 days

### 2.12.1 Harvest, purification, and transduction of HSC lin- with lentiviral vectors

Mice donor bone were harvested, and bone marrow (BM) flashed out. BM pulled from all donors' bones was then subjected to lineage negative depletion kit (Miltenyibiotec, 130-110-470). Purified HSCs lin neg cells were seeded in $4x10^6$/ml in StemSpan™ SFEM (StemCell Technologies, 09600) supplemented with 2% FCS, 1% P/S and following cytokines: 100ng/ml mSCF (mouse stem cell factor) (StemCell Technologies, 78064), 100ng/ml mFlt3 (StemCell Technologies, 78011) and 25ng/ml hTPO (StemCell Technologies, 78210.1). A total of 220µl of LV-SSPNewTagI-IIGAA or 195µl of LV-GAA-Native viral production was added to appropriate wells with purified lin neg to total of MOI 150. In addition, samples of untransduced HSC were also seeded at $4x10^6$ /ml as a control group. Subsequently, HSCs were subjected to spinoculation at 500g for 1 hr and transferred for 24hrs incubation at 37°C and 5% $CO_2$. After 24 hrs incubation cells were washed twice with 5ml PBS, span down at 300g for 5 mins and counted so that even number of cells was transplanted across all constructs/ animals. Even number of cells was resuspended in 100µl of cold PBS and processed to I.V. injection.

### 2.11.2 Liquid culture

24 hrs post-transduction cells were harvested, span down at 300g for 5mins, washed twice in PBS and resuspend in 1ml of PBS. Subsequently HSCs were counted and $100x10^5$ cells per each condition (transduced/untransduced) were seeded into 100µl of StemSpan supplemented as previously in Section 2.10.1. and transferred to an incubator at 37°C and 5%$CO_2$ for further 7 days. After this period cells were harvested, washed with PBS and pelleted down. Pellets were resuspended in 50ul of PBS containing

2mM $MgCl_2$ buffer (Sigma-Aldrich, 1374248-1G) and treated for 30 mins at 37°C with Benzonase® Nuclease enzyme (Sigma-Aldrich, E1014-25KU) to digest any residual viral particles/genetic material possibly still present on cells surfaces. Consequently, cells were resuspended in 1ml of PBS, further split into to 500 µl samples in 1.5 Eppendorf tubes with 1st tube being processed for DNA purification to determine VCN (Section 2.4.1) and the 2nd for GAA activity (Section 2.4.2). Eppendorf's were then span down at 300 g for 5 mins and pellet transfer to -80°C.

### 2.12.3 Colony Forming Units – CFU

In addition to Liquid culture also, CFUs were seeded 24hrs post transduction. For each condition (transduced/untransduced) $3.3 \times 10^3$ cells were added into 3ml of MethoCult™ (Stem Cell Technologies, 04434), well resuspended and distributed 1ml per each 3.5cm petri-dish (Thermo Scientific™ Nunc™ Cell Culture/Petri Dishes, 10441013), for a total of 3 dishes/condition. Petri-dishes were then transferred to 37°C and 5% $CO_2$ incubator for 14 days. Further, CFUs were inspected and counted. Dishes from each condition were then pulled together into a 15ml tube and washed with 5ml PBS, span down at 300g for 5 mins and cell pellets treated with Benzonase® Nuclease enzyme as previously described in 2.10.2. Cell pellets were divided evenly for DNA extraction to determine VCN (chapter 2.4.1) and also for GAA assessment (chapter 2.4.2) and stored at -80°C.

### 2.12.4 Busulfan regimen

Prior to transplant of allogenic transduced HSCs, recipient mice went under Busulfan (Merck, B2635-25G) regimen to myeloablation of the bone marrow (BM) compartment. Mice were subjected to 25mg/kg IP injection daily for 4 days prior to transplant. Twenty-five mg of Busulfan were carefully weighted and dissolved in 1ml of acetone (Merck, 179124-500ML). Consequently, busulfan/acetone solution was further diluted by adding 9ml of peanut oil (Merck, P2144-250ML). This very viscous solution was then gently but toughly mixed by pipetting it up and down until formation of a uniform solution. Busulfan solution was prepared afresh daily.

**2.12.5 Functional test**

To assess gene therapy outcomes and determine muscle correction, a rotarod (Basile Microprocessor Controlled RotaRod Treadmills, 57601) test took place. Each animal was placed on a slowly spinning rod for 5 mins to allow the animal to adjust to the speed and a new environment. After acclimatisation, the rod started gradually increasing its velocity so that each animal was forced to keep moving progressively faster and also coordinate its position (front and back) in relation to the spinning rod. Upon reaching a muscle exhaustion, examined animal triggered a ratchet which in turns stopped a timer. The timer was activated as the rotarod transitioned from acclimatisation phase into acceleration period, allowing for precise time entry for each animal separately. Each animal was a subject to 5 consecutive runs with 3 minutes break between each individual run. Only the last three runs were used for analysis.

**2.12.6 Flow cytometry**

To determine lineages distribution, in harvested animals' bone marrow, two panels of cellular markers was used. Panel A consisted of APC fluorochrome antibody (Invitrogen, 17-0452-82) against B220 cellular marker and FITC combine with anti CD3e marker (Invitrogen, 11-0031-82). Panel B comprised of APC anti CD11b (BioLegend, 101208) and FITC with anti Gr1 cellular marker (Invitrogen, 11-5931-82). Each antibody was diluted to 1:100 using Magnetic-activated cell sorting buffer (MACS), containing 0.5% of BSA, 2mM EDTA (Sigma-Aldric, ED2SS-50G) and PBS. Resuspended in MACS buffer $10x10^5$ BM cells were mixed with panel of antibodies. For control samples, $10x10^5$ BM cells in 100 ul of MACs buffer were mixed with 1μl of antibody mounted with FITC or APC fluorochrome. All samples containing cells were then subject to an incubation for 25min at RT with panel A, B, and individual controls. Consequently, stained samples were span down for 5 mins at 300g, supernatant removed and 150μl of MACS buffer used to wash cells. Samples were then span again at 300g for 5 mins and supernatant removed. Cell pellets were resuspended in 150μl of MACS buffer. Final step consisted of addition of 50μl of 4% formaldehyde (Thermo Scientific™, 28908) to each sample to fix the cells. After gently mixing stained BM cells with fixative, samples were processed with flow

cytometry using BD™ LSR II flow cytometry cell analyser (BD Bioscience, LSR II). Analysis of flow cytometry data was processed using FlowJo™ v10.8 software. Each sample was gated for general population using side scatter (SSC) vs forward scatter (FSC), follow by single evens population only using forward scatter- area (FSC-A) vs forward scatter hight (FSC-H). Lastly, single cell population was gated on particular cellular marker vs FCS-A plot. Any event showing at least $log10^2$ particular cellular marker brightness intensity was counted as a positive event for the marker.

## 2.13 Statistics

Details of specific statistical tests used during this study can be found in the relevant figure legend. All statistical analyses were conducted using GraphPad Prism software version 8.3.0 (https://www.graphpad.com/scientific-software/prism/).

# Chapter 3   Results

## 3.1 Modification of GAA at N-terminus

The native protein of acid alpha glucosidase (GAA) remains intracellularly produced, folded, glycosylated and trafficked into lysosomes where it serves to break down glycogen. Less than 10% of the total natural GAA protein is secreted outside the 'producer' cell (Braulke & Bonifacino, 2009). In order to force the producer cells to secret recombinant human GAA (rhGAA) and in turn to improve secreted rhGAA protein to be taken up by target cells i.e., myofibres, the N-terminus of GAA enzyme can be modified to accommodate all the challenges of secretion, uptake and lysosomal targeting. While the N-terminus of proteins can be modified to facilitate different goals, inadvertent compromise of activity and/or stability might jeopardise gene therapy outcomes. In this chapter I will explore different options to optimise rhGAA protein secretion and increase its uptake in myotubes.

## 3.1.1 Muscle Homing peptide – 'NewTag'

Pompe disease is not effectively treated by ERT leaving many individuals without a full correction of the disease (Rohrbach & Clarke, 2007). Certainly one main problem is the insufficient plasma membrane expression of M6PR in muscle tissue (Martin-Touaux et al., 2002) and BBB in CNS (Puertollano & Raben, 2018). To improve efficacy of GAA uptake in skeletal muscles, we designed recombinant forms of GAA enzyme which were engineered to express muscle homing peptides at N-terminus.

The original size of the GAA protein consists of 952 amino acids (aa), in which initial 45 aa at N-terminus serve as a native signal peptide, which is crucial for the enzyme to be translated within the ER. After translation, the signal peptide is cleaved off and removed from the main structure of the enzyme (Braulke & Bonifacino, 2009). Based on previous studies, 69 amino acids were removed from the N-terminus leaving GAA enzyme from

70<sup>th</sup> amino acid onward (GAA70) with high enzymatic activity (Maga et al., 2013). So, removing 69 aa retains GAA activity and allows N-terminal modification to be undertaken without compromising enzymatic activity.

I first tested a muscle homing peptide- called 'NewTag' in this study, which has been described previously in literature (Gao et al., 2014) to have high affinity to skeletal muscles. This is a 12 amino acid long peptide, RRQPPRSISSHP, which was reverse translated into a human codon optimised DNA sequence. Since the mechanism of action for this peptide is unknown, alongside a construct bearing 1 motif of NewTag I (Figure 3.1.1 A) I also cloned one carrying a tandem version of NewTag (NewTag I-II) (Figure 3.1.1 B). As NewTag motif plasma counterpart remains unknown, the lucidity behind this exercise was to test potential benefit of two muscle homing peptide domains over a single domain to potentially enhance binding affinity and stabilize the protein and its plasma membrane counterpart/receptor interaction.

To determine the impact of NewTag domains on GAA protein, the enzymatic activity was measured in pellets of transduced K562 cells (Figure 3.1.1 C) and in cell medium (Figure 3.1.1 D) and normalized by VCN. The GAA activity in cell pellets shown significant difference between the GAA Native construct and NewTagI and NewTagI-II (\*\*$p<0.01$, and \*$p<0.05$ respectively, +/-SEM, n=3). In addition, significantly higher GAA activity in cell pellets was observed between NewTag I and NewTag I-II in the favour of two domain construct (\*\*$p<0.01$, +/-SEM, n=3). In cell medium, NewTag I GAA construct shown higher GAA activity to NewTag I-II GAA, however the difference was not significant ($p=0.075$, +/-SEM, n=3). In addition, both constructs showed significantly higher GAA activity compered to GAA native construct (\*\*$p<0.01$, +/-SEM, n=3).

We have concluded that NewTagI-GAA and NewTagI-IIGAA secreted in media shows similar enzymatic activity and therefore both designs were taken for further analysis including intracellular uptake and lysosomal targeting.
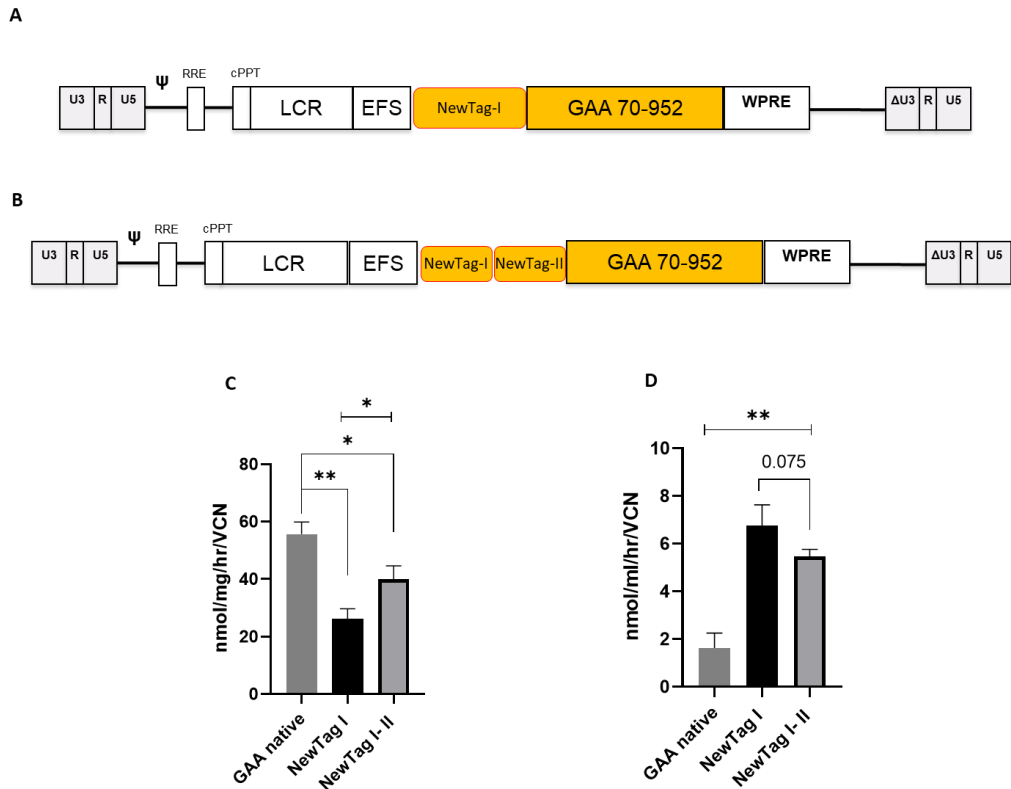
***Figure 3.1.1 Muscle homing peptide – structure and impact on GAA enzymatic activity.***
*Map of lentiviral vectors showing the truncated GAA70 with **A)** 1 motif of NewTag
(NewTag I) or **B)** with the tandem motif of NewTag (NewTag I-II) . Enzymatic activity of
transduced K562 cells with Native GAA (baseline control), NewTag I or NewTag I-II
normalized by VCN of individual samples, **C)** in cell pellets and **D)** in medium . (\*\*p<0.01,
\*p<0.05, +/-SEM, n=3). One-way ANOVA with Tukey's post hoc test was used to alanyse
experimental groups.*

### 3.1.1.1 Altered amino acid lengths of GAA

In 2005 ZyStor Therapeutics, Inc. (Milwaukee, WI 53226-4838 (US)) filed a patent testing
several different truncations to GAA protein (U.S. patent No. PCT/US2005/004286). GAA
protein showed various levels of its activity regardless of the truncation size. In addition,
our initial GAA70 truncation size originated from this study. I have decided to pick
additional two truncations to test GAA enzymatic activity in hope to increase enzymatic
activity, if possible.

To preserve as much as possible the native structure of the enzyme, I decided to test truncated rhGAA forms: GAA55 and GAA50, where 15 or 20 amino acids were added from the native GAA sequence.

However, GAA50 and GAA55 did not improve enzymatic activity neither in transduced K562 cell pellets (Figure 3.1.2 A) nor in medium (Figure 3.1.2 B), and caused the opposite results, a fall of activity below the one measured for the GAA70 truncated construct. In particular, the GAA enzymatic activity was decreased by 43% and 58% in cell pellets and by 16% and 48% in medium for GAA55 and GAA50, respectively, compared to GAA70. For this reason, we continued our study utilising the GAA70 form.
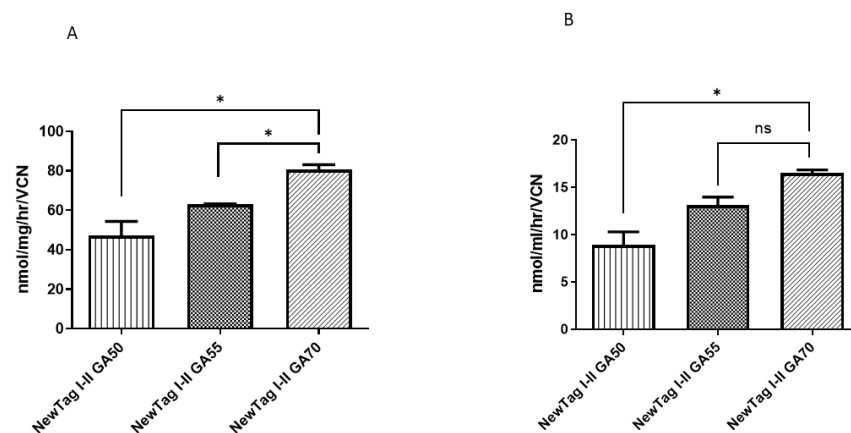


***Figure 3.1.2. Impact of truncation lengths on GAA activity.*** *GAA activity normalized by VCN in **A)** cell pellets and **B)** medium of K562 cells transduced with NewTag-I-II GAA50, NewTag-I-II GAA55 or NewTag-I-II GAA70. All conditions were transduced at MOI 10. (\*p<0.05, ns= not significant, SEM +/-, n= 2, One-way ANOVA with Tukey's post hoc test).*

### 3.1.1.2. Increasing GAA protein stability trough amino acid linkers

Further attempts to improve GAA activity and protein-protein interactions brought me to study protein linkers. There are many studies exploring the importance of protein linkers, where they serve to improve protein quality, and to produce stable protein–

protein complexes and modify an interface between two or more components of recombinant proteins (Chen et al., 2013; Chichili et al., 2013; Klein et al., 2014). Protein linkers can provide with firm support between main protein and any added modification or they can be more flexible to help to accommodate extra structures without interfering with the main protein functional domains (Chen et al., 2013).

To improve further truncated GAA70 protein and increase its catabolic activity, a protein linker was added between the modified N-terminus and the main body of GAA70. In this study, two sizes of linkers composed of neutral amino acids were used, precisely 3 and 13 amino acid combination, GAP and NLNGLAAVLKGAP, respectively. Both NewTag I and NewTag I-II modified GAA constructs were cloned in lentiviral vectors with 3 or 13 aa linkers upstream of GAA (Figure 3.1.3 A) and used to transduce K562 cells.

GAA enzymatic activity in K562 cell pellets showed that neither 3 nor 13 aa linker improved potency of rhGAA, regardless of the NewTag motif examined. By contrast, the longer the linker, the greater was the decrease in enzymatic activity observed in comparison to a construct without any linker (Figure 3.1.3 B) (**$p<0.01$, SEM+/-, n=3). Similarly, in medium from transduced K562 cells GAA activity of the secreted enzymes reduced with increased size of amino acid linker (Figure 3.1.3 C) (**$p<0.01$, SEM+/-, n=3). According to these data, the absence of linker between the NewTag and GAA70 promoted greater GAA enzymatic activity.
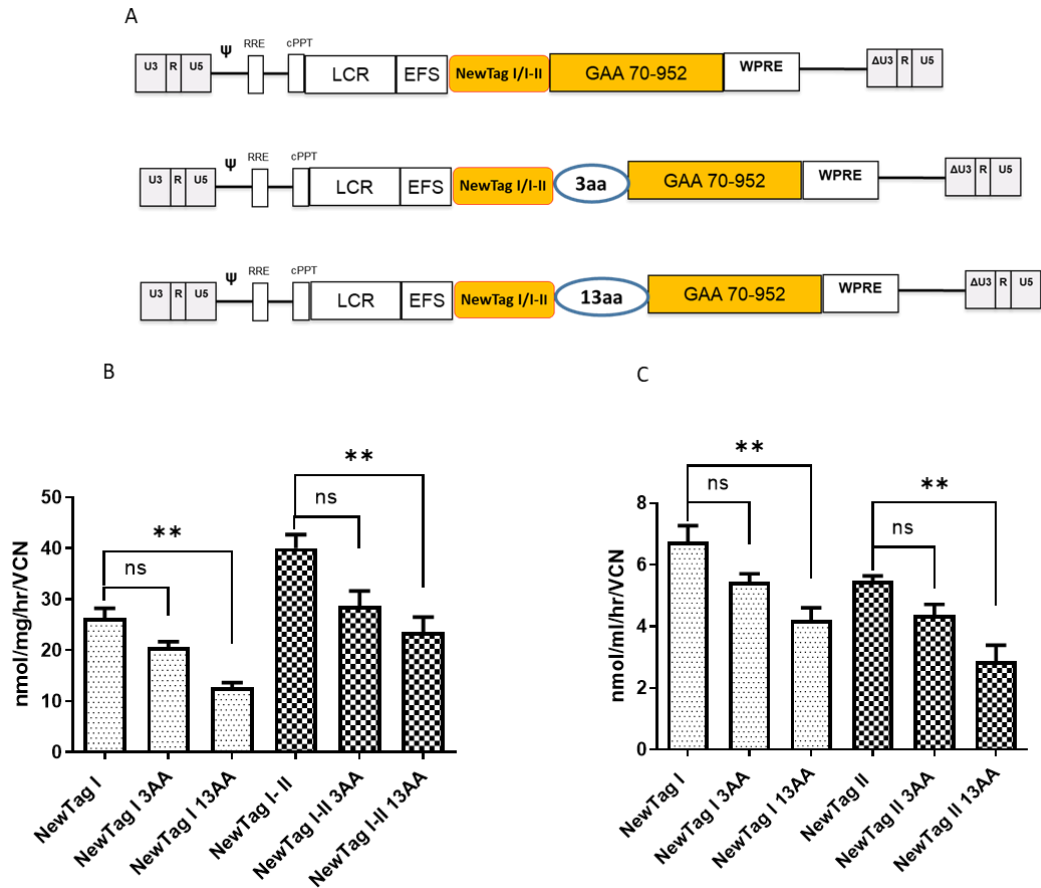
***Figure 3.1.3. Increasing GAA protein stability by applying different number of amino acid linker.*** *Linear structure of lentiviral vectors and rhGAA components with (**A, top**) no aa linker, (**A, middle**) 3 aa or (**A, bottom**) 13 aa linker between the N-terminal modification and GAA70. Addition of amino acid linkers in the transgene showed decrease in GAA activity for both NewTag I/ I-II and both in **B)** cell pellet and **C)** medium (\*\*p<0.01, +/-SEM, n=3, One-way ANOVA with Tukey's post hoc test). All values of GAA activity are normalised by VCN from individual samples.*

## 3.1.2 Uptake study of NewTag modified GAA by confocal microscopy

For easy detection in confocal fluorescent microscopy, I used a previously cloned lentiviral vector containing hGAAmCherry construct (Figure 3.1.4 A) that has mCherry reporter in frame with GAA at C-terminus. The use of GAAmCherry fused protein more readily allowed the study of GAA localization upon uptake in treated cells.
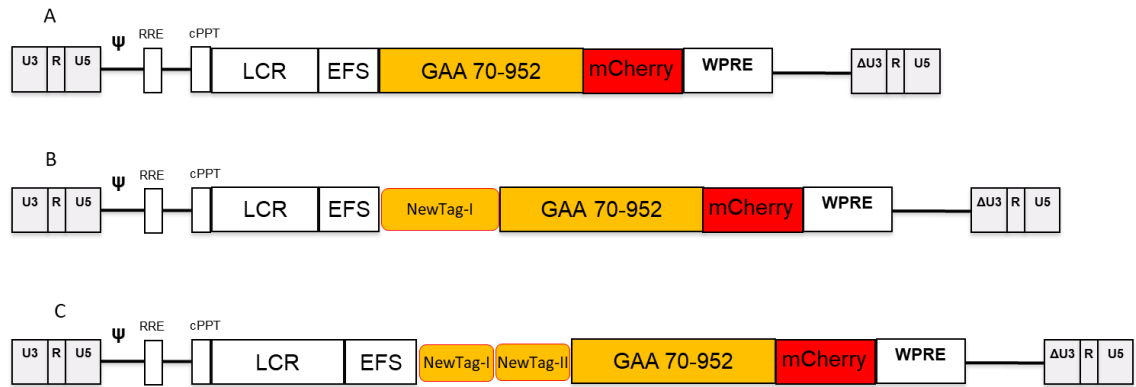
***Figure 3.1.4. Structural organisation of viral vectors containing fused hGAAmCherry.
A)*** *Original structure truncated GAA containing mCherry structure. Following **B)** with addition of NewTag I and **C)** NewTag I-II modifications on lentiviral vectors containing hGAAmCherry.*

Both NewTag I- and NewTag I-II- hGAAmCherry were cloned into pCCL- lentiviral backbone to produce lentivirus particles and stable K562 cell lines expressing these two constructs (Figure 3.1.4 B and C). Seven days after transduction K562 cells were FACS sorted to acquire a stable line expressing high mCherry signal. Medium from these cells were harvested, concentrated using Pierce™ Protein Concentrators and tested for GAA molar concentration. These were used as source of GAA to test in uptake assays.

### 3.1.2.1. Localization study upon uptake of NewTagI-GAA

NewTagI-GAAmCherry concentrated medium was used to treat differentiated H2B 2B4 myotubes for 24 hrs. Live cell imaging in confocal microscopy was then used to detect mCherry within the myofibers. Lysotracker was added, to visualise live lysosomes, and M6P competitive antagonist was used to prevent and exclude any uptake through CI-M6PR. The M6P inhibitor was added 1 hr prior to the addition of 1μM NewTagI-GAAmCherry concentrated medium, and both were incubated with cells for the duration of 24 hr. We hypothesised that any mCherry signal within cells treated with M6P competitive antagonist would represent recombinant GAA utilising an alternative route of entry to CI-M6PR.

To localise GAAmCherry, three different channels were recorded from each imaged field: bright field, red and green signals, which allowed recognition of cell borders, GAAmCherry and lysosomes, respectively. Figure 3.1.5 illustrates H2B 2B4 differentiated myotubes in two conditions of M6P inhibition: no inhibitory M6P molecule added (A-F) and with the M6P inhibition (A'-F'). In samples treated with M6P followed by exposure to NewTagI-GAAmCherry, the mCherry signal seems less abundant within treated myotubes (C'). In contrast, H2B 2B4 without M6P inhibition and only treated with NewTagI-GAAmCherry, the mCherry signal within the myotubes appears more plentiful. This would suggest that internalisation and lysosomal targeting of NewTagI-GAAmCherry molecules relies substantially on CI-M6PR. However, some mCherry signal is still visualised under M6P inhibitory conditions, suggesting that NewTagI-GAAmCherry is internalised and reaches lysosomes in a potentially CI-M6PR independent manner.

The Image J software with a micro-Java script allowed me to overlay mCherry channel/signal with lysosomal green channel/signal from the confocal microscopy pictures. This created a composite picture, green (lysosomes), red (mCherry) and where these two colours overlayed a yellow colour appeared, suggesting co-localisation of mCherry inside lysosome. Confocal microscopy taken from H2B 2B4 myotubes +/- M6P and NewTagI-GAAmCherry were analysed, and composite pictures evaluated. Further, quantification of the yellow signal/channel was conducted (Figure 3.1.5 G) revealing that the co-localisation signal was more abundant in H2B 2B4 treated only with NewTagI-GAAmCherry (Figure 3.1.5 E), and less rich in myotubes in M6P inhibitory condition follow by exposure to NewTagI-GAAmCherry (Figure 3.1.5 E'). This would support the previous observation with mCherry signal evaluation (Figure C and C') that lysosomal transport of NewTagI-GAAmCherry depends substantially on CI-M6PR. However, the positive co-localisation observed under M6P inhibitory condition (Figure 3.1.5 E') allows us to presume that NewTagI-GAAmCherry lysosomal internalisation might be yet facilitated by alternative routes to CI-M6PR. Furthermore, due to an error in processing dapi stained nuclei of myofibers were not acquired.

In addition, H2B 2B4 myofibers were very sensitive to movement (transport), pH and temperature change. Despite best efforts to preserve optimal conditions while imaging H2B 2B4 myofibers i.e., heat confocal microscopy chamber, some of the cells underwent very fast contraction losing its original elongated shape and becoming irregular shaped

globular structures with all their contents compressed together. Red channel illustrating mCherry signal was condensed in a small space giving the impression of GAAmCherry protein, in presence of its antagonist, to form extracellular aggregates. However, in fully elongated H2B 2B4 cells mCherry signal seems more spread across the cell with less sense of aggregates being formed as in case of collapsed H2B 2B4 myofibers. As myofibers tend to form elongated, branching out structures, with many parts of one myofiber intertwining with neighbouring myofibers, in the case of sudden contractions of a myofiber underlying another elongated H2B 2B4 would give a false impression of mCherry aggregates presence. Therefore, only fully elongated H2B 2B4 myofibers were taken under considerations and analysed for potential GAAmCherry alternative route of uptake in a presence of its antagonist. Quantification of mCherry signal overlapped with lysosomal lysotracker signal was possible only in elongated myofibers. In contrast, a quantification of the overlapped signal was impossible to conduct in contracted myofibers and therefore it was omitted from analysis.

E



F



E'



F'



**co-localisation**

**co-localisation**

G
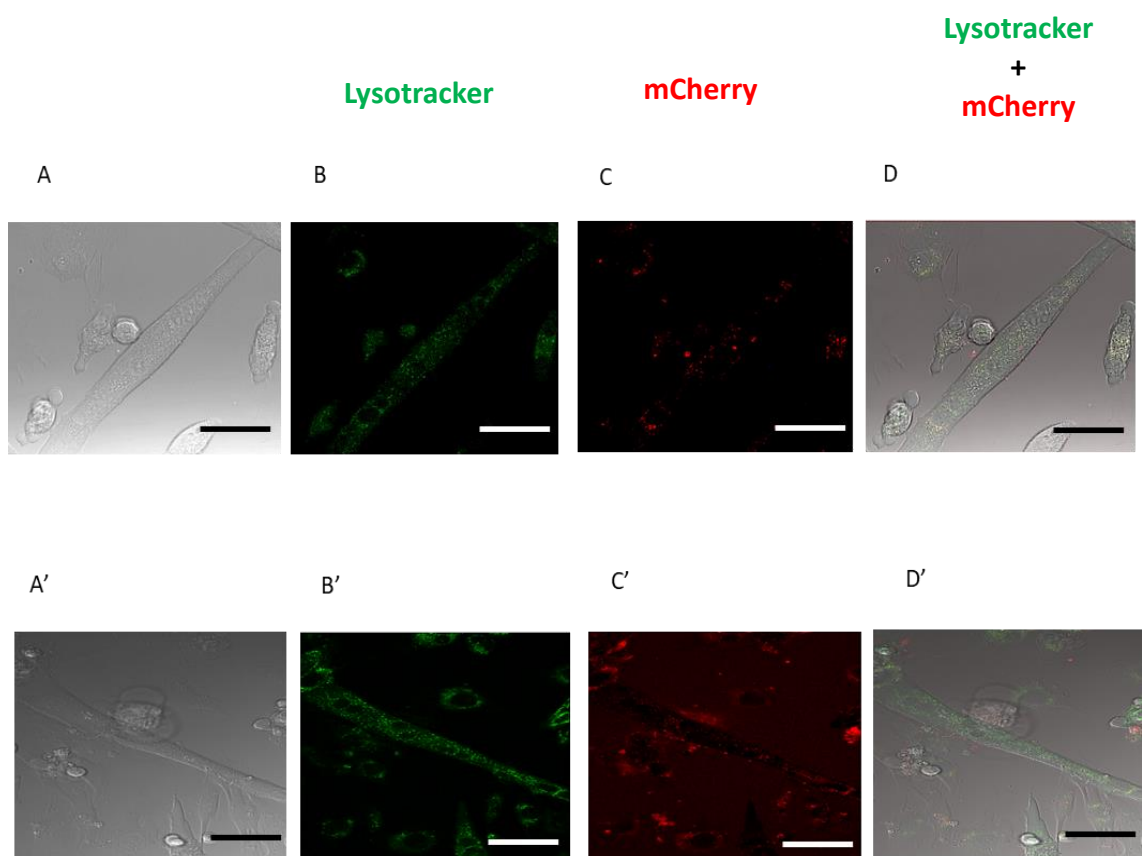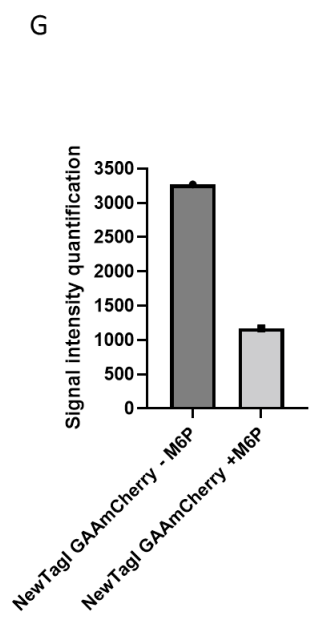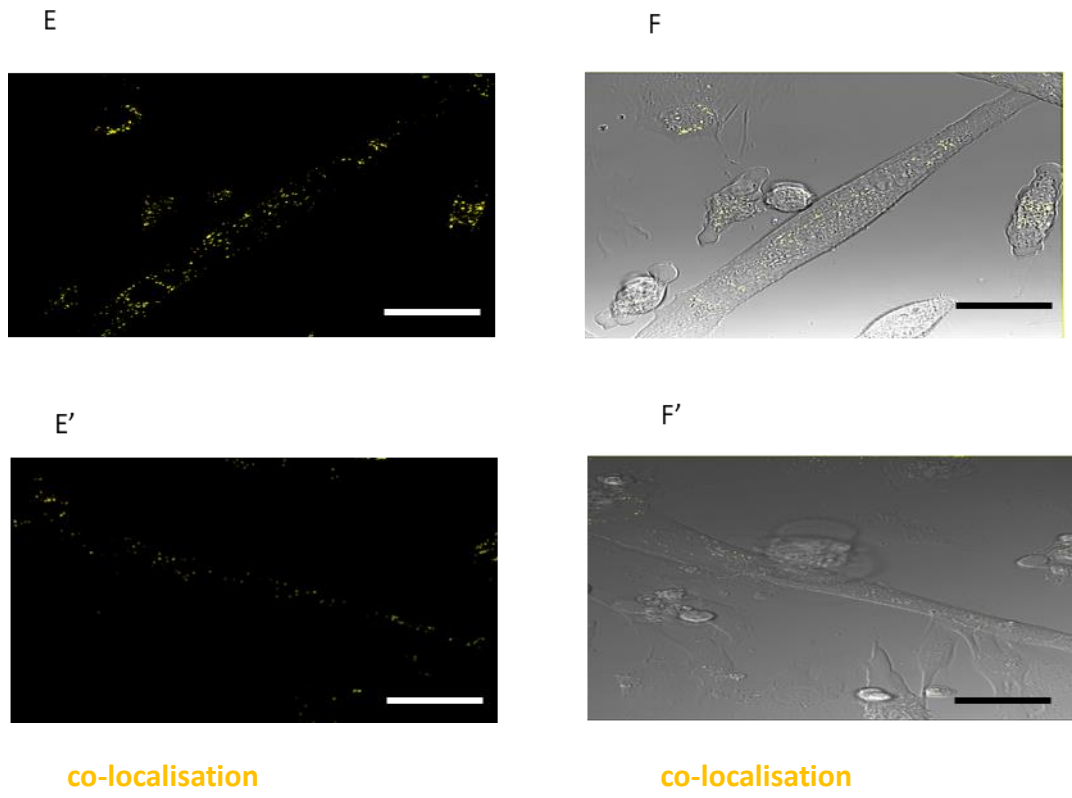


*Figure 3.1.5. Localisation study of NewTagI-GAAmCherry upon uptake in cells.* *H2K 2B4 myotubes were exposed to 1µM of NewTagI-GAAmCherry in absence (A-F) or presence (A'-F') of 5mM of M6PR competitive antagonist (M6P) for 24hrs. Lysotracker was added*

*at 5nM concentration into cell culture 1 hr prior imaging. Each sequence of images consists of a bright field acquisition (A, A') to indicate cellular plasma membranes borders, lysotracker signal (B, B') to show lysosomal structures and mCherry signal (C, C'). Composite pictures (D, D') were made by combining bright field, lysotracker and mCherry signal together. Co-localisation of mCherry in lysosomes was presented as a yellow colour, which came from the overlapping of green and red signal (E, E'). Co-localisation signal was merged with bright field acquisition to authenticate intracellular origin (F, F'). Quantification of co-localization signal (G) (n=1). Live imaging was performed in confocal microscopy heated chamber at 37°C. Scale bar set for 50µm.*

### 3.1.2.2. Affinity comparison of NewTagI-GAA versus NewTagI-II-GAA

An additional confocal microscopy experiment was designed to study the implication of 1 domain (NewTagI) vs. tandem domain (NewTagI- II) in H2K 2B4 myotubes uptake. As in the previous chapter 3.2.1. the contracted H2B 2B4 myofibers were removed, where appropriate, from analysis as quantification of the overlapped signal mCherry and lysosomal lysotracker was impossible to conduct.

Myotubes were exposed to 1µM either of NewTagI-GAAmCherry or NewTagI-II GAAmCherry concentrated medium. On this occasion all conditions were treated with 5mM M6P inhibitor for 24 hrs. As previously described, several images including bright field, mCherry and lysotracker acquisitions were carried out for each imaged field and micro-Javar script on Image J was used to highlight the yellow signal arisen from the overlapping of mCherry signal with lysosomes in green. Both variants, NewTagI (Figure 3.1.6 A-F) and NewTagI-II (Figure 3.1.6 A'-F') showed the ability of facilitating the entry of GAAmCherry protein into treated myotubes (Figure 3.1.6 C-C'). Both modifications at the N-terminus of GAA construct are potential candidates to target its cellular counterpart, however the co-localization analysis in treated myotubes (Figure 3.1.6 E and E') indicated that the tandem structure might have a stronger affinity to the binding partner producing more the co-localisation signal (Figure 3.1.6 G). Due to an error in processing Dapi stained nuclei of myofibers were not acquired.

Lysotracker · mCherry · Lysotracker + mCherry

A  B  C  D

A'  B'  C'  D'

E  F

E'  F'
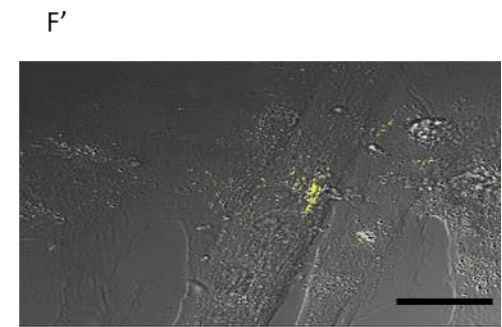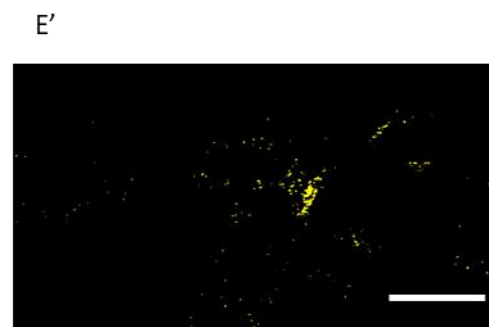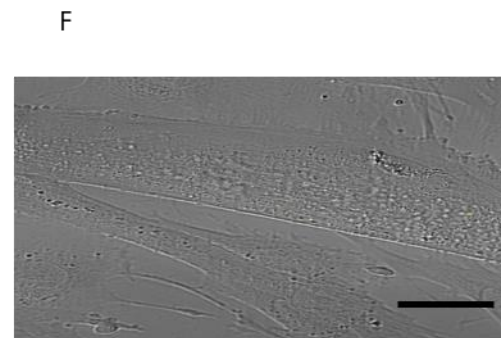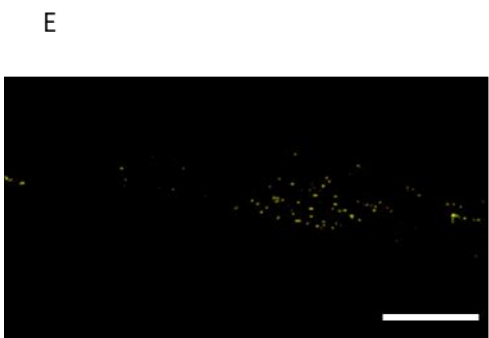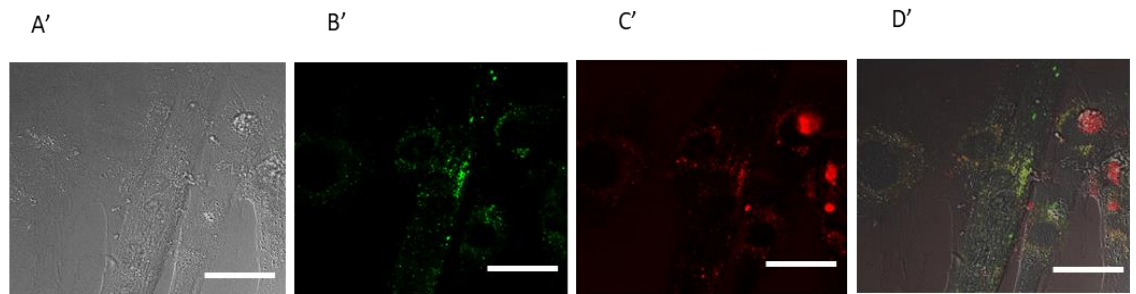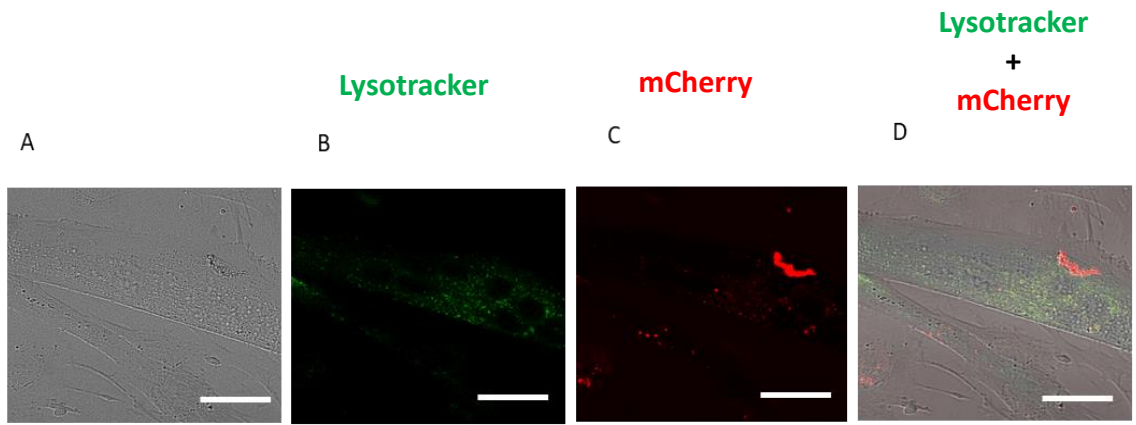
co-localisation · co-localisation
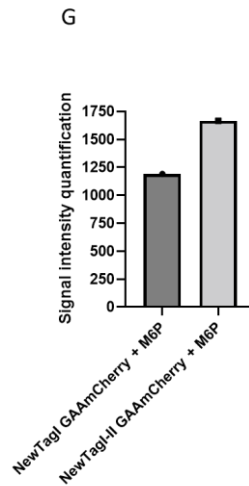
*Figure 3.1.6. Localization of NewTagI vs. NewTagI-II tandem GAAmCherry upon uptake in M6P treated myotubes. H2K 2B4 were exposed to 1µM of NewTagI-GAAmCherry (A-F) or NewTagI-II-GAAmCherry (A'-F') concentrated medium in presence of 5mM M6P inhibitor for the duration of 24 hr. Lysotracker for lysosomal staining was added to cell culture 1hr prior to cell live imaging. For both muscle homing peptide constructs bright field (A, A'), lysotracker (B, B') and mCherry (C, C') acquisitions were taken. Composite was generated by merging the acquisitions together (D, D'), while using micro-Javar script the co-localisation of mCherry and lysosomes was shown in yellow colour (E, E'). To determine if the co-localisation was within plasma membrane, the yellow signal was merged with bright field acquisition (F, F'). Quantification of co-localization signal (G). Live imaging was undertaken in confocal microscopy heated chamber at 37ºC. Scale bar set for 50µm.*

### 3.1.3. Improve the secretion signal peptide

N-terminal signal sequence facilitates targeting of new secretory proteins to the endoplasmic reticulum (ER) in a signal recognition particle (SRP)-dependent manner (Mercier et al., 2017). The signal peptide is recognised by a signal peptidase and cleaved to create the complete protein that can be trafficked through the Golgi network before being secreted from the cell via secretory pathways (Gritti, 2011). There are many naturally occurring small signal (also called secretion) peptides, located at the N-

terminus of a protein, that allow for release of secretory proteins in turn (Pera et al., 2001). The native form of GAA protein does not possess a secretory signal peptide as this protein is not a secretory protein (Roig-Zamboni et al., 2017). In order to enforce the secretion of transgenic GAA, a secretory signal was incorporated into the N-terminus replacing a part of the deleted native amino acids. The N-terminus of NewTag modified GAA was equipped with the Insulin growth factor secretion signal of IGF-II (Pera et al., 2001). This secretion peptide provided a moderated increase of GAA protein being secreted into cell medium and a slight reduction of both GAA isoforms, 110kDa and 76 kDa, within cells compared to native GAA, as illustrated on Western blot probed for GAA protein (Figure 3.1.7 A and B, respectively). To note, an accumulation of 110kDa was detectable intracellularly in cells that carried the N terminal modification of GAA, which might decrease therapeutic efficacy (Figure 3.1.7 B).

Guler-Gane *et al.* in his study compared several naturally occurring signal peptides with a silico synthetic peptide (SSP) (with amino acid sequence consisting of MWWRLWWLLLLLLLLWPMVWA) and their ability to support protein's secretion. SSP showed nearly 3 folds increase in protein secretion compared to alternative signal peptides used. Following this study I decided to replace IGF II signal peptide by SSP (Guler-Gane et al., 2016). K562 cells transduced with SSP-NewTagI-II-GAA showed a dramatic decrease in GAA enzymatic activity in cell pellets compared to constructs carrying the IGF-II signal peptide or the native GAA (Figure 3.1.7 C) which also correlated to a significant decrease in both GAA isoforms in the cell lysate as showed by western blot (Figure 3.1.7 D). On the other hand, the GAA activity in medium of those cells showed a dramatic increase by 230% comparing to constructs bearing the same muscle homing peptide upstream of GAA but with IGF II signal peptide as a secretion peptide (Figure 3.1.7 E) ($p<0.001$, +/-SEM, n=3). Also, the Western blot analysis of those medium showed a substantial increase of GAA protein concentration when SSP was used as secretion peptide compared to IGF-II or the native signal peptide (Figure 3.1.7 F). In particular, SSP-NewTagI-IIGAA construct was 3-fold higher secreted than the equivalent GAA construct with IGF-II signal peptide (Figure 3.1.7 G). These results show that SSP is a powerful alternative to IGF-II secretory peptide to be used in combination with muscle

homing peptide and rhGAA in order to reduce accumulation of rhGAA protein in the factory cells and further increasing its secretion and bioavailability for cross-correction.
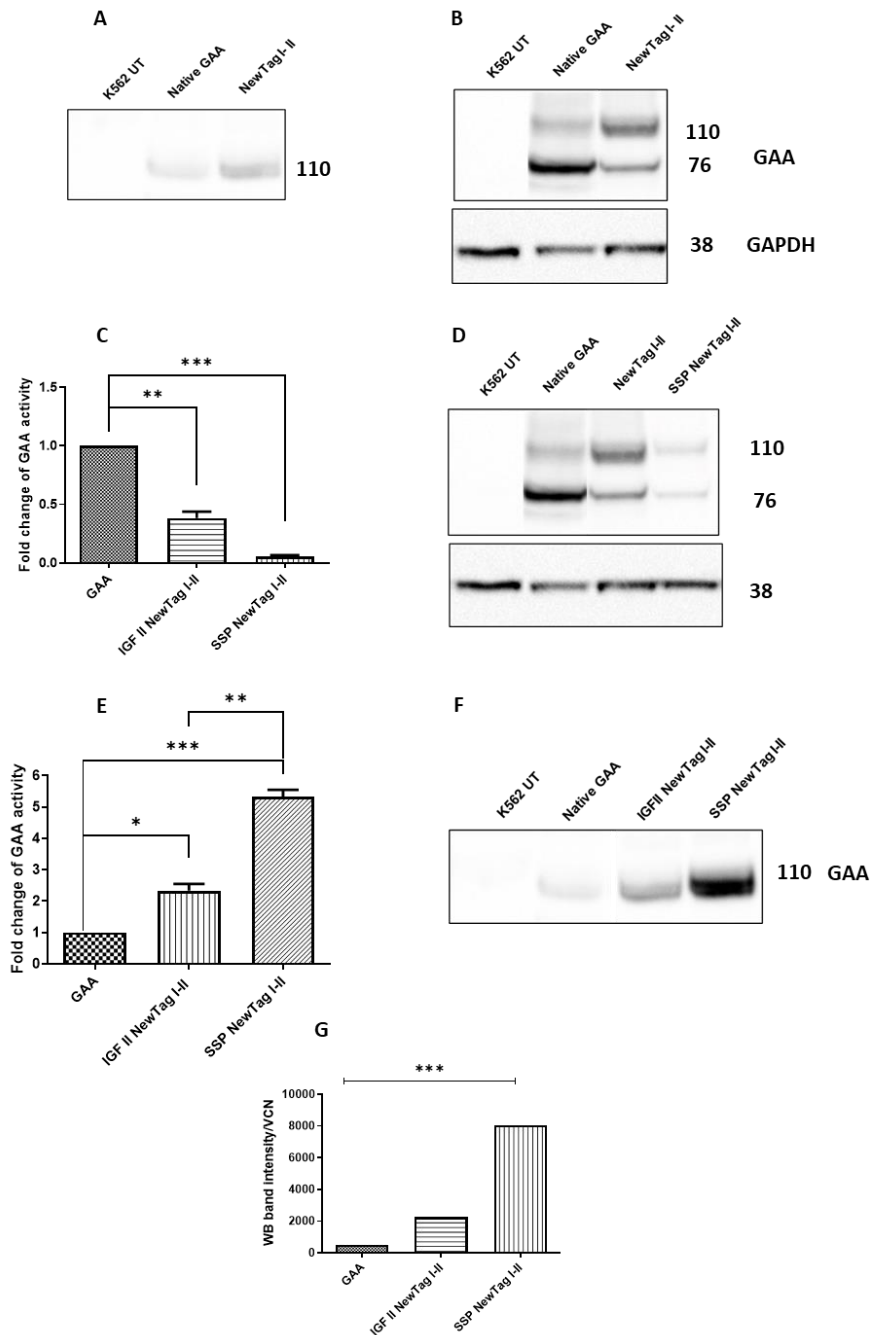


***Figure 3.1.7. Evaluation of secretory peptide SSP on GAA recombinant protein. A)*** *Western blot for hGAA of cell medium from K562 cell lines bearing different variants of*

*NewTag I-II GAA with IGF-II signal peptide or native rhGAA. The 110KDa isoform is observed. **B)** Intracellular rhGAA protein in transduced K562 cells compared with untreated K562, internal isoforms of GAA at 110 and 76 kDa are observed. GAPDH was used as loading control. **C)** Fold change of GAA activity of individual constructs containing either IGF-II or SSP signal peptide upstream of NewTag modified GAA compared to native GAA (\*\* p<0.01, \*\*\*p<0.001, respectively, SEM+/-, n=3). **D)** Western blot analysis of transduced K562 cell pellets with different GAA constructs showing internal isoforms of GAA at 110 and 76 kDa. GAPDH was used as loading control. **E)** GAA activity in cell medium from transduced K562 cells with different GAA constructs. Construct containing SSP modification showed an increase in GAA activity of 230%, compared to the IGF II equivalent GAA constructs (\*\*p<0.01, +/-SEM, n=3). **F)** Anti-GAA immunoblot of medium samples showing secreted GAA precursor at 110 kDa. **G)** Quantification of the western blot bands from immunoblotted medium samples presented in F showed significant difference in band intensity between SSPNewTag IIGAA, IGF bearing construct and Native GAA (\*\*\*p<0.001, n=1). SSP bearing sample showed an increase of GAA protein level of 3-fold compared to the equivalent construct bearing IGF II secretion peptide. Fold change of GAA activity and Western blot intensity quantification were normalised by VCN from individual samples. One-way ANOVA with Tukey's post hoc test.*

### 3.1.4 β-galactosidase control assay

Inside lysosomes GAA enzyme targets exclusively glycogen, however in laboratory settings detection and determination of glycogen degradation is difficult to trace and instead fluorogenic 4-Methylumbelliferyl-α-D-glucopyranoside substrate is used to determine GAA level and its enzymatic activity (Chamoles et al., 2004). In lysosomes β-galactosidase (β-gal) catalyses degradation of galactosylceramide to ceramide and galactose with its optimal working pH range between 3.5 to 4.5 (Martarello et al., 2019). To detect β-gal enzymatic activity *in vitro* a synthetic fluorogenic substrate 4-methylumbelliferyl-**β**-D galactopyranoside is used (Chiu & Watson, 2017). Since both synthetic substrates once cleaved generate 4MU, the same fluorometer assay can be used to detect the signal generated by the activity of both enzymes. It is not reported that β-gal can break down GAA's 4MU substrate, however since both enzymes are

lysosomal hydrolases and laboratory conditions vary from optimal *in vivo* environment i.e., pH, temperature etc., allowing some aberrations in substrate recognition and its processing by examined enzymes (Garcia-Contreras et al., 2012). Running the two enzymatic activities in parallel has been used as quality control for the cell and lysosomal lysing procedure. To exclude any possibilities that the lack or modulation of GAA enzymatic activity recorded might be in fact due to variations in the lysosomal protein extraction, β-gal was tested alongside GAA to verify the accuracy of GAA enzymatic activity assay in this study. Untreated and transduced (LV SSPNewTagI-IIGAA) K562 cells along with untransduced 293T cell line were tested for GAA and β-gal activity.

Cell medium and pellets from both K562 and 293T were subject to 4-methylumbelliferyl-**β**-D galactopyranoside or 4-Methylumbelliferyl- **α**-D-glucopyranoside. Following 1 hr incubation at 37°C, enzymatic activity was read and extrapolated from a standard curve of know concentration of 4MU. Untransduced K562 cell pellets showed a low native level of enzymatic activity for both enzymes. Therefore, I used 293T cells as a positive control for the biochemical reactions since these cells are rich in lysosomal enzymes (Trivedi et al., 2020).

The analysis of enzymatic activity, presented in Figure 3.1.8, showed that the level of β-gal enzymatic activity in cell pellets and medium of transduced K562 cells remained unchanged indicating that the increase of GAA activity observed upon cell manipulation was truly due to the modulation in enzyme expression and secretion.
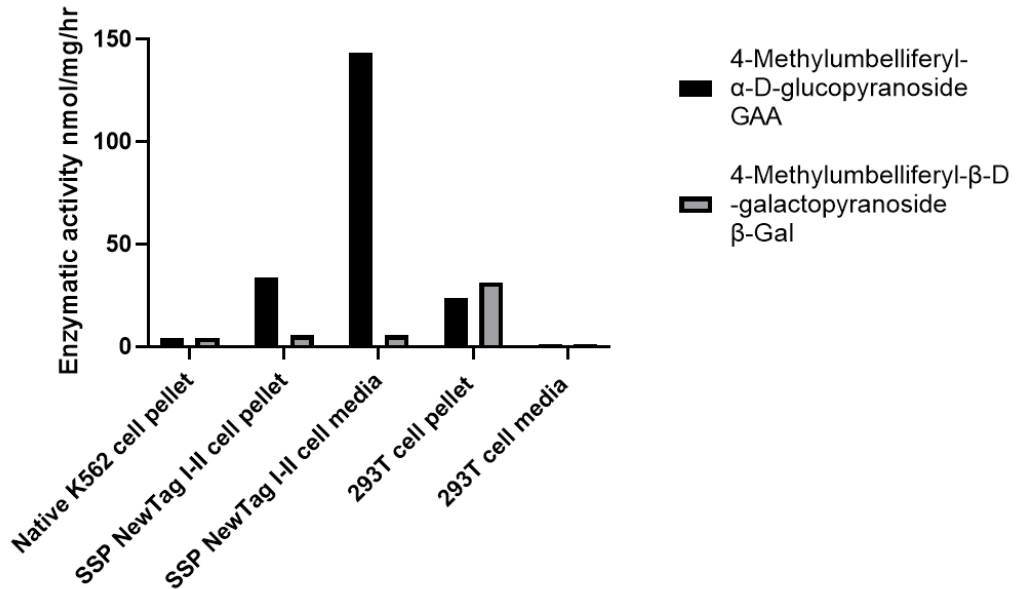
***Figure 3.1.8. Analysis of enzymatic activities GAA and β-Gal.*** *Cell pellets and medium from untransduced and transduced K562 and untransduced 293T cells were used to establish enzymatic activities of GAA and β-gal using their specific synthetic substrates 4-Methylumbelliferyl-α-D-glucopyranoside and 4-methylumbelliferyl-**β**-D galactopyranoside, respectively (n=1).*

### 3.1.5 Discussion chapter

This chapter of work details the development and optimisation of a novel design of the N-terminus of rhGAA construct. The modification of N terminus of this protein was necessary to equip GAA with a novel muscle homing peptide.

Due to difficulties in processing imagingof live myotubes using confocal microscopy in evaluation of the SSPNewTagI-IIGAA

Based on previous work done by ZyStor Therapeutics, Inc. (Milwaukee, WI 53226-4838 (US)), several different truncated GAA protein forms were tested to determine their GAA enzymatic activity while modifying the N-terminus. Truncation or any other modification to the primary structure can affect the enzymatic activity of the modified protein. As observed in my study extension of native amino acid structure was counterproductive

as it decreased the enzymatic activity. Similarly, addition of protein linkers to potentiate protein-protein interaction (NewTag-its receptor counterpart) showed diminished effect to the overall GAA activity and this approach was abandoned. It appeared that different configuration of amino acids would be perhaps more advantageous, however exploration of this field would greatly extend beyond the time frame of this project and was not further evaluated.

In addition to GAA enzymatic activity, an uptake of GAAmCherry fused protein utilising NewTagI/I-II, +/- M6P inhibition, into H2B 2B4 myofibers was studied. An overall suggestion that NewTag allows GAA protein for lysosomal targeting was drawn. However, the H2B 2B4 myofibers structures were very difficult to preserve intact under confocal microscopy imaging, potentially obscuring the full picture of the uptake, and therefore it was decided to omit this approach in the further investigation in favour of other methods to determine NewTagI-IIGAA lysosomal uptake and glycogen reduction.

In the uptake experiments (Chapter 3.1.2) only co-localization signal was quantified as a primary indication of the NewTag motifs' ability to facilitate lysosomal targeting. The red channel representing mCherry was omitted from quantification as live imaging proven to be very challenging, with many myotubes rapid contraction impacting a quality of pictures taken. We decided that a single channel was insufficient to determine a cellular internalisation and only co-localisation arisen from two channels overlapping was sufficient to draw conclusions.

Furthermore, final part of this chapter was focused on signal peptides, structures lying at the N-terminus of every known protein (Teufel et al., 2022). These peptides have a fundamental significance in defining a fate of proteins. In a study by Guler-Gane *et al.* explored several different secretory proteins supporting export of a protein extracellularly. By incorporating one of the signal peptides a substantial build-up of rhGAA protein in transduced K562 cells was resolved. Firstly, more rhGAA protein was secreted and was available to deliver cross-correction and secondarily less of the potentially toxic accumulation of produced rhGAA protein was observed in transduced cells.

It is also very important to mention about putative immunoreactivity towards NewTagI-IIGAA protein as a potentially unknown foreign object introduced into blood circulation. Particularly CRIM negative patients suffer from production of neutralising antibodies against ERT using rhGAA as a respond to foreign/exogenous proteins. The patient's immunological system mounts a humoral responds against ERT damping down the therapeutic effects and correction of PD (Al Khallaf et al., 2013). One concern with our approach is that there may be an immune response to our construct especially in light of adding a signal peptide or linker to the rhGAA. An immune response of this nature is of paramount importance as it might jeopardise all efforts related to PD correction. However, in case of autologous transplant of transduced HSCs an induced tolerance to rhGAA is likely to be acquired (Cencioni et al., 2022).

All progeny of transduced HSCs, including antigen presenting cells (APC), carry a copy of the transgene and produce the recombinant protein. After transplanting transduced HSCs their progeny i.e., APC migrate to lymphoid organ and specifically to the thymus which allows for the negative selection of developing autoreactive T cells thereby generating central tolerance. Peripherally transgene product-derived peptides are presented to developing naïve CD4$^+$ cells (Bowen et al., 2002; Khanna et al., 2000). This leads to removal and anergy of recombinant protein-specific effector T cells (Teff) and to induction of CD4$^+$CD25$^+$FoxP3$^+$ T regulatory cells (Treg) (Luth et al., 2008; Tang & Bluestone, 2008). Through synergy with APCs, induced Treg censor CD8$^+$ T cells (CD8$^+$ are mediators of adaptive responses) and Teff. Furthermore, inhibition of Teff leads to an absence of T help and therefore suppression of B cell responses that can mount antibody formation followed by cytotoxic T lymphocyte (CTL) reactions (Mingozzi et al., 2003; Sakaguchi et al., 2008; Skaggs et al., 2008). We predict that APC-dependent induced tolerance to NewTagI-IIGAA (and all its components including SSP) in PD patients will prevent adaptive responses against any components of design in this study rhGAA that might be later used in *ex vivo* GT. However, this would need to be further assessed *in vivo* using appropriate PD mouse model.

# Chapter 3.2 Analysis of gene expression, glycosylation and uptake of N-terminus modified rhGAA *in vitro*.

### 3.2.1 Posttranslational glycosylation modifications of rhGAA

Posttranslational modifications of proteins consist of crucial processes providing modified peptides with correct folding orientation, methylation, glycosylation, and therefore providing all means for a protein to be fully functional (Braulke & Bonifacino, 2009; Moreland et al., 2005). These processes are undertaken in strictly controlled and defined cellular circumstances and any deviation from the cellular 'schedule' might lead to skipping or missing of some vital parts of posttranslational modification resulting in the production of a dysfunctional protein (Wang et al., 2014). N-linked oligosaccharides are a glycosylation modification that is particularly important in protein uptake by target cells through CI-M6PR. N-linked oligosaccharides higher in mannoses are attached to 7 specific amino acids on GAA chain and are then further phosphorylated in Golgi forming mannose-6 phosphate residues (Moreland et al., 2005). Since high efficiency and speed of secretion of the rhGAA proteins might affect posttranslational modifications, we instigated if the introduction of a strong signal peptide such as SSP could interfere with the correct pattern of glycosylation of GAA.

Glycosylation patterns were studied using purified and concentrated glycoproteins by HiTrap ConA 4B columns from medium of transduced K562 cells. Two specific glycosidases recognizing N-linked oligosaccharides were employed to determine the glycosylation pattern in native GAA and GAA carrying IGF-II or SSP NewTagI-II tag. In particular, PNGase F recognises and removes all N-linked oligosaccharides from glycoproteins while Endoglycosidase H cleaves only within the chitobiose core of high mannose and hybrid oligosaccharides of N-linked glycoproteins (Figure 3.2.1).
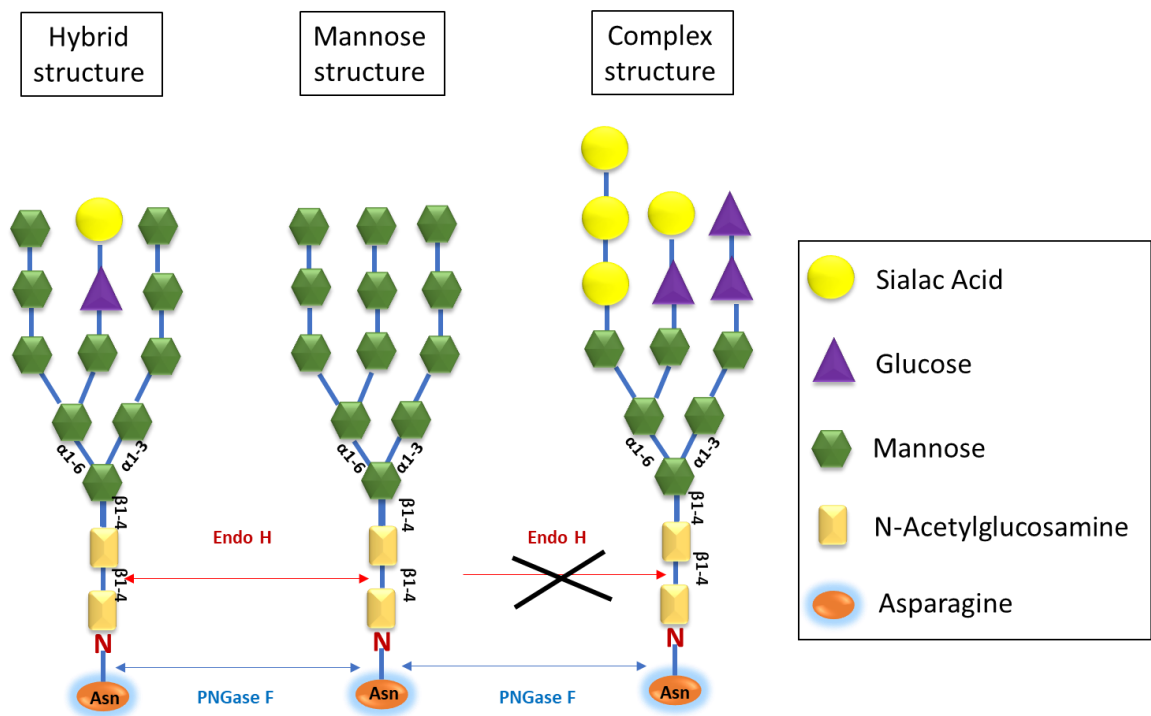
**Figure 3.2.1. Cleavage sites of PNGase and Endo H glycosidases.** *N-linked oligosaccharides recognised by PNGase and Endo H. PNGase (arrows in blue) removes all N-linked oligosaccharides from the structure of glycoproteins and Endo H (arrows in red) cleaves off only structures of high mannose and/or hybrids, if mannose is in the minority Endo H dissociate from glycoprotein without removal of the structure.*

Glycoproteins including secreted rhGAA were processed with PNGase or EndoH and then analysed by Western blotting with anti-human GAA. In addition, purified glycoproteins from tested recombinant GAA forms were also blotted uncleaved to determine molecular weights prior to enzymatic digestions (Figure 3.2.2 A). Immunoblotting with anti hGAA antibody unveiled that SSP or IGF-II in GAA constructs resembled an identical pattern of bands and band sizes compared to the native GAA both in untreated (Figure 3.2.2 A) or treated samples with Endo H (Figure 3.2.2 B) or PNGase F (Figure 3.2.2 C). We hypothesised based on these findings that SSP peptide did not alter the N-linked oligosaccharide pattern in rhGAA.
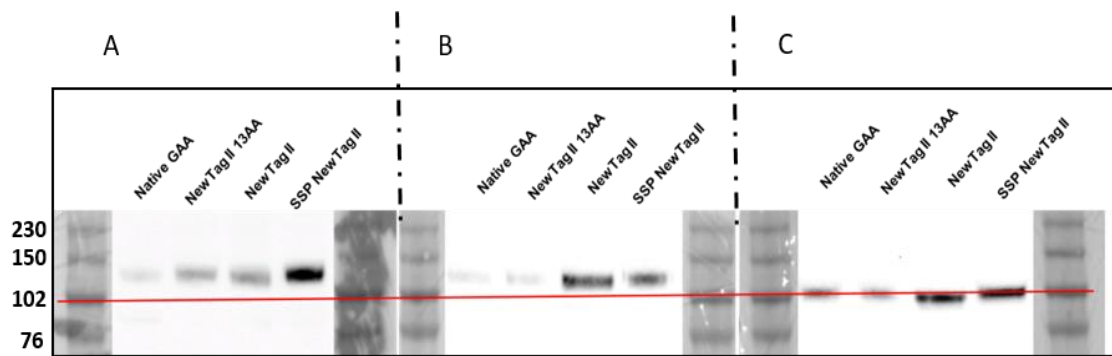
*Figure 3.2.2. Examination of N-linked oligosaccharide pattern in rhGAA by Western blotting. A)* *Untreated glycoproteins purified from medium of K562 transduced with native hGAA or different variants of NewtagI-II GAA were blot for hGAA. Bands exhibited identical molecular weight of 110kDa across all modified rhGAA and native GAA. B)* *Samples treated with Endo H or C)* *treated with PNGase F showed identical distribution and molecular size without any visible aberration when compared to native GAA. Glycoproteins were loaded in each line of Western blot gel at different amount for each blot. Red line across is placed on 102 kDa molecular size based on full range molecular markers.*

## 3.2.2 Gene expression analysis of GAA constructs under different gene promoters

In addition to translational and posttranslational modifications, critical to the success of gene therapy and gene transfer is also the transcription of gene of interest. Since majority of viral promoters have been removed from viral vectors for safety reasons, transgene cassette relies entirely on internal transgene promoters for stable transgene expression. Variety of transgene promoters have been used with some various levels of success (Nieuwenhuis et al., 2021; Nieuwenhuis et al., 2023; Zheng & Baum, 2008). Virus-derived promoter i.e., the spleen focus-forming virus (SFFV) or myeloproliferative sarcoma virus promoters (MND) are relatively strong promoters, providing high gene expression, however viral promoters raise safety concerns to switch on adjacent oncogenes through the presence of enhancer sequences (Zhou et al., 2016) Therefore, mammalian ubiquitously expressed promoters i.e., acting elongation factor 1α promoter (EFS 1α) or/and 3-phosphoglycerate kinase (PGK) have been preferentially

chosen to power transgene expression due to their lack of trans activating potential. However although safe, the mammalian promoters have diminished gene expression potential to that of virally-derived promoters , therefore tissue specific promoter or/and tandem of promoters are tested to deliver best gene therapy outcomes (Colella et al., 2019; Rein et al., 2003).

To determine a level of GAA transgene expression under locus control region of human β-globin gene (β-LCR) I chose a human erythroleukemia cell line K562 to work with as a main cell line prototype (Andersson et al., 1979). The β-LCR specifically boosts the expression of the β-globin or specific heterologous promoters in erythroid progenitors and could significantly enhance gene expression from haematopoietic cells (Montiel-Equihua et al., 2012). This K562 line, however modified/immortalised *in vitro*, can provide important information about the degree of enhancement to GAA transgene expression I could expect *in vivo* using β-LCR in erythroid progeny developed from transduced HSC. In addition, K562 cell lines share similarities with erythroid blood lineages, was also used to determine rhGAA protein posttranslational modification, enzymatic activity and secretion levels that were hoped to obtain in *in vivo* study.

### 3.2.2.1 Comparison of PGK and LCR-EFS1a promoters

Promoters within viral vector designs are used to ensure expression of a transgene in the host genome. In this section I will be comparing two regulatory cassettes, β-LCR-EFS 1α and PGK, to determine the GAA expression in K562, an erythroid like cell line, and 293T cells.

Vital components of the locus control region of human β-globin gene (β-LCR) were combined with the constitutively acting elongation factor 1α promoter (EFS 1α). It was previously shown that βLCR can upregulate EFS 1α activity particularly in erythroid cells but did not enhance its action in lymphoid and myeloid lineages (Montiel-Equihua et al., 2012). βLCR-EFS 1α is a relatively large fragment measuring about 3kb in length, which might have negative implications on LV packaging and there affect LV titre and because of the enhancer activity of the LCR, this may affect transcription of neighbouring genes. On the other hand, the PGK promoter, reaching about 500bp in length, is upstream of

the 3-phosphoglycerate (*Pgk-1*) ubiquitously expressed gene, which is involved in glycolytic pathways and therefore universally present and constitutively expressed in all somatic cells (Sawado et al., 2003) and has been widely used in clinical trials thereby offering a high level of reassurance on safety (Sevin & Deiva, 2021).

Lentiviral vectors carrying either βLCR-EFS 1α or PGK promoter upstream of *GAA* gene were used for the transduction of K562 or 293T cells. After 5 days post-transduction, cells were spun down and pellets were divided for genomic DNA purification and GAA enzymatic activity. Media samples were also kept to assay GAA enzymatic activity. VCN was assessed from purified genomic DNA samples using qPCR. GAA enzymatic activity in cell pellets and media were assessed by biochemical assay as indirect method to measure GAA expression. In K562 cells, GAA activity in cell pellets and medium normalised by VCN showed a significant advantage of βLCR-EFS 1α over PGK promoter for each tested construct (Figure 3.2.3 A and B). While, in HEK 293T cells the differences between these two promoters were less pronounced than in K562 cells, however the βLCR-EFS 1α promoter showed greater activity compared to PGK promoter for several different constructs (Figure 3.2.3 C and D).
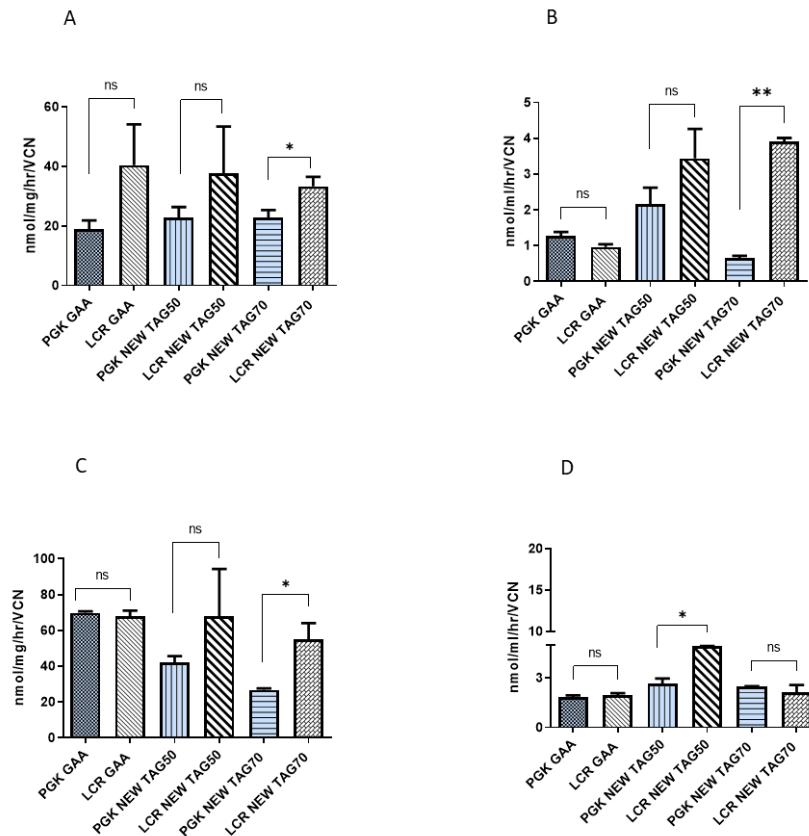
*Figure 3.2.3. PGK vs. β-LCR EFS1α comparison in K562 and HEK 293T. 5 days post transduction with different forms of rhGAA, K562 cell line were harvested, A) cell pellets and B) medium samples were processed form GAA activity assay. In addition, HEK 293T cell lines were also transduced and harvested for GAA activity in C) cell pellet and D) medium samples. GAA activity in cell pellets and medium samples were subject to normalisation by VCN. Both K562 and HEK 293T cell line were transduced with MOI of 10. GAA activity in cell pellets were expressed and plotted as nmol/mg/hr and medium samples as nmol/ml/hr using GraphPad Prism. (\*p<0.05, +/-SEM, n=2, one-way ANOVA with Tukey's post hoc test).*

## 3.2.2.2. Transcriptional analysis of integrated lentivectors carrying LCR-EFS 1α promoter.

Gene expression of different constructs was conducted in comparison to the native GAA construct to exclude any issue at transcriptional level, which would then reflect in a difference in protein translation and therefore enzymatic activity in the several constructs. Gene expression was performed by RT-qPCR platform by using delta-delta

Ct method, also known as the $2^{-\Delta\Delta Ct}$ method, which allowed the calculation of the relative fold gene expression of modified forms of GAA vs. native GAA transgene expression (Figure 3.2.4). According to these data the different modifications in GAA sequence do not interfere with the transcription rate and all constructs showed similar level of transgene expression when compared to the expression of native *GAA* sequence.
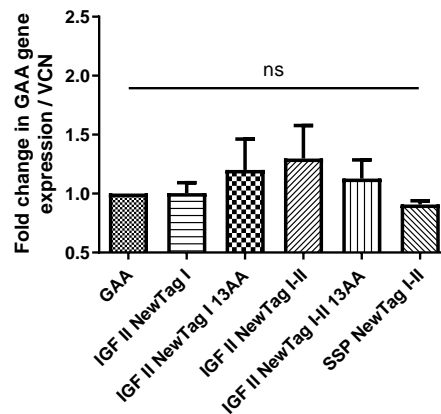


***Figure 3.2.4. GAA gene expression analysis.*** *Values of gene expression were analysed using $2^{-\Delta\Delta Ct}$ equation and graphed as a fold change vs. native GAA. Acquired data showed various gene expression patterns across human GAA construct bearing several NewTag modifications and IGF II with SSPNewTag I-IIGAA resembling closest expression level to native GAA. (p=ns, +/-SEM, n=3, one-way ANOVA with Tukey's post hoc test).*

### 3.2.3. *In vitro* characterisation of N-terminus modifications of rhGAA

The preceding part of this chapters was focused to design the most robust and effective GAA construct which equipped the enzyme with promising muscle homing peptide and secretion peptide. In this part of chapter 3.2, N-terminal modifications of GAA recombinant protein will be tested for the uptake by target cells and reduction of glycogen stored in lysosomes in a relevant cell model.

### 3.2.3.1 Generation of C2C12 *Gaa* knockout cell line

H2K 2B4 myoblast were used to assess the ability of muscle homing peptide NewTagI/I-II to deliver human GAAmCherry to lysosomes. However, this cell line expresses endogenous GAA enzyme and in addition is very difficult to handle in culture. In order to test the specific uptake, it was decided to create a GAA knock-out myoblasts using C2C12 murine cell line. The creation of a cellular model of Pompe disease was ideal to test rhGAA uptake and consecutively glycogen modulation upon uptake in knock-out cells.

Out of 10 clones collected and tested using ICE, two clones revealed adequate mutations in *Gaa* gene (Figure 3.2.5 A) leading to complete cessation of GAA enzymatic activity compared to C2C12 wild type cells (Figure 3.2.5 D). Nevertheless, Western blot showed surprisingly presence of GAA protein at ~ 102 kDa in both clones, however processed isotype at 76 kDa was missing compared to wild type C2C12 samples (Figure 3.2.5 C). The antibody used to detect GAA protein targeted exon 3 according to manufacture (Abcam, ab137068), however its exact motif was a propriety of Abcam PLC and not available. This information suggest that the mutations introduced in exon 2 did not propagate to exon 3, as amino acid composition in exon 3 is most likely unchanged allowing aforementioned antibody to bind their target. Mutations introduced to exon 2 and their exact impact on mRNA splicing events leading to formation of immature GAA proteins are difficult to explore at this point of study (Abramowicz & Gos, 2019; Anna & Monika, 2018; Baeza-Centurion et al., 2020). Nevertheless, WB data showed that mature form of rhGAA protein formation is prohibited in mutated clones, suggestion that acquired amino acid aberration prevented the protein in question from maturations.  In addition, a lack of basal level of GAA activity in mutated clones compared to wild type permitted us to use this cellular model of Pompe disease in further analyses of cross correction.

**A**

| | | | | | |
|---|---|---|---|---|---|
| C2C12 wild type | PAM | TGGTTGCTTCTT | | | |
| C2C12 sgRNA 2 clone 2 | PAM | TGGTGT--TTCTT | | | |
| C2C12 sgRNA 2 clone 3 | PAM | TGGTTGTTTCTT | | | |

**B**

| | PAM | W | L | L | L |
|---|---|---|---|---|---|
| C2C12 wild type | PAM | W | L | L | L |
| C2C12 sgRNA 2 clone 2 | PAM | W | C | - | |
| C2C12 sgRNA 2 clone 3 | PAM | W | L | F | L |

**C**



GAA 102 / 76

GAPDH 37

**D**

| GAA AC | | |
|---|---|---|
| | | nmol/mg/hr |
| C2C12 GAA KO sgRNA 2 | clone 2 | 0.30 |
| | clone 3 | 0.20 |
| | C2C12 Wild type | 50.89 |
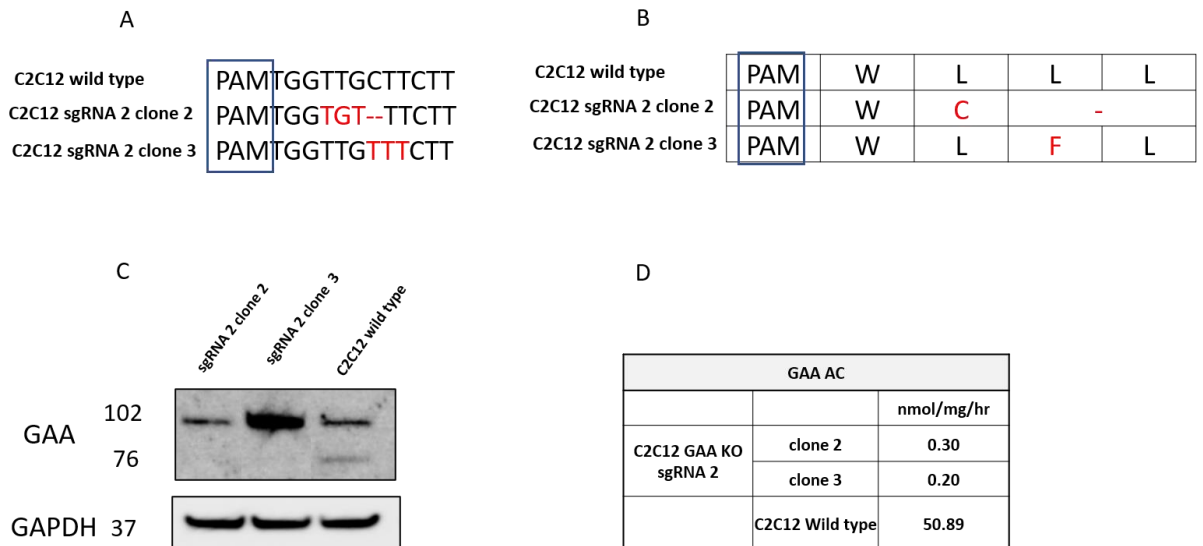
***Figure 3.2.5. Validation of Gaa -/- C2C12 cell line.*** *Sanger sequence revealed mutations in **A middle lane)** clone 2 (in red), in particular the substitution of two base pairs -G for T and T for G- and the loss of a C base pair causing reading frame shift. Also clone 3 showed one substitution **A bottom lane)** (in red) of T to C close to the PAM sequence. **B)** The mutations in both clones altered the amino acids sequence of GAA. **C)** Western blot analysis displayed residual juvenile form of GAA protein in both clone 2 and clone 3. **D)** Enzymatic activity from lysed cells of both clones showed a total lack of GAA catabolic activity compared to WT C2C12 cells indicating full enzymatic knockout in both clones.*

### 3.2.3.1.1 Assessment of CI-M6PR on C2C12 ko cells surface

To use *Gaa -/-* C2C12 as *in vitro* model of PD, I investigated if any of the off-target mutations carried by sgRNA 2 had potential impact on CI-M6PR expression in this cell line. To avoid any bias in GAA uptake using *Gaa -/-* C2C12 cells, an analysis of CI-M6PR cellular expression was conducted using western blotting and confocal microscopy. For this purpose, $2 \times 10^5$ cells of both WT and *Gaa -/-* C2C12 line were seeded in 6 well plate and differentiated in myofibers for 5 days. Then cells were harvested and proteins from plasma membranes isolated. Only plasma membrane fractions were collected to eliminate any contamination of M6PR originated from cellular compartments. Proteins from plasma membrane fractions were immunoblotted using anti CI-M6PR alongside

with anti Na⁺K⁺ATPase pump that served as an internal control due to its ubiquitous plasma membrane expression (Clausen et al., 2017). Protein content was determined by BSA assay and 20μg of total proteins from each sample was analysed. Based on the immunoblotting findings, *Gaa -/-* C2C12 cells did not show any variability in CI-M6PR protein level compared to WT C2C12 in two independent experiments (Figure 3.2.6 A).

In addition to immunoblotting, *Gaa -/-* and WT C2C12 cells were imaged at confocal microscopy to determine the distribution of CI-M6PR on their plasma membranes. Cells were seeded for both conditions in confocal microscopy chamber slides and upon differentiation fixed. Permeabilization step was not performed to prevent signal that might arose from internal M6PR molecules. The cellular distribution of CI-M6PR in both samples clone 2 (Figure 3.2.6 B i-iv) and WT (Figure 3.2.6 B i'-iv') showed a similar punctuated pattern (Figure 4.6 B i and i'), providing enough confidence to use the *Gaa -/-* C2C12 line in future experiments for GAA uptake.
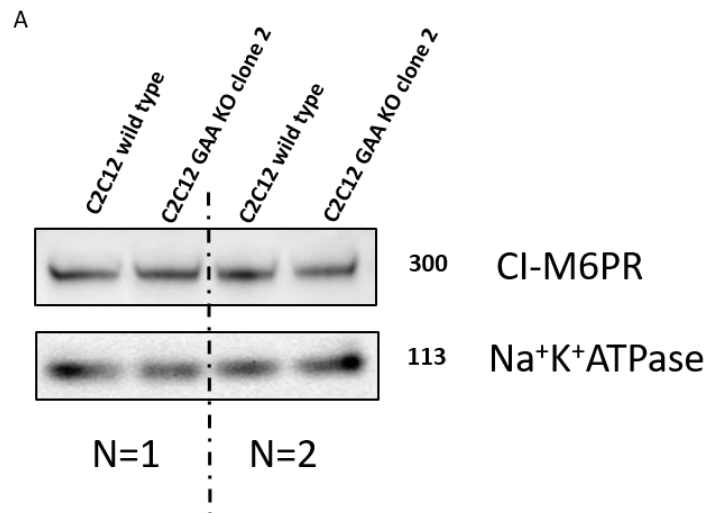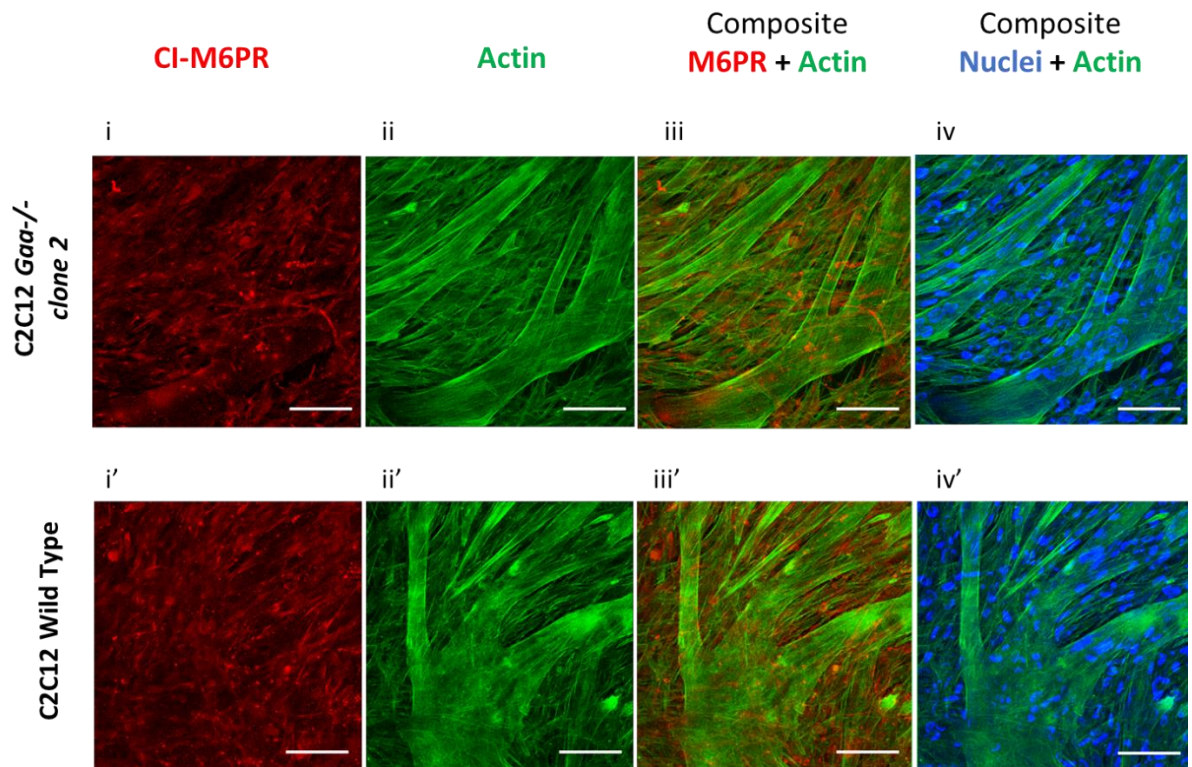
*Figure 3.2.6. Study of Cation Independent Mannose-6-Phosphate Receptor expression on plasma membranes of Gaa -/- and WT C2C12 cells. A) Western blot analysis of plasma membrane bound proteins for CI-M6PR from Gaa -/- and WT C2C12 cells (n=2). Blots showed no significant difference in CI-M6PR between the two cell lines. Na+K+ATPase was blot as loading control. B) Confocal microscopy of Gaa -/- (i-iii) and WT (i'-iii') C2C12 cells stained with anti CI-M6PR (i and i'), Phalloidin F-actin (ii and ii'). Composite images of CI-M6PR and actin filaments was performed to show CI-M6PR distribution on cells (iii and iii') (n=1). Dapi staining for nuclei visualisation and composite image with Actin staining (iv and iv'). Dapi staining was compromised by incompatible with the dye washing buffer used, leading to only partial staining of the nuclei. Scale bar set for 50µm.*

### 3.2.4 rhGAA uptake in transwell co-culture

Previous study of GAAmCherry uptake into myofibers, using confocal microscopy, proven particularly difficult to conduct and conclude. Myofibers exposed to external

conditions, losing its elongated shapes, contracting, and concentrating their intracellular compartments into irregular shaped globular structures proven arduous and challenging to investigate. Therefore, alternative methods of GAA protein uptake and lysosomal targeting were employed.

*Gaa -/-* C2C12 myoblasts were used to determine a level of uptake of GAA recombinant proteins when treated with different rhGAA constructs in a transwell co-culture model (Figure 3.2.7 A). *Gaa -/-* C2C12 cells were seeded at $2x10^5$ cells per transwell in duplicates. Upon differentiation into myofibers, K562 cell lines carrying different rhGAA constructs with similar values of VCN (3±0.3) were placed in the inserts of the transwell plate. Each insert had micro pores of 0.4μm in diameter allowing only molecules like enzymes to migrate freely through the membrane but not the physical contact between different cell types in co-culture. To determine which of the N-terminal modification containing muscle homing peptides exhibits a truly M6P-independent cellular entry, one well from each pair of wells per individual rhGAA had added of 5mM M6P that is CI-M6PR competitive antagonist. After 24 hrs co-culture transwells were disassembled, C2C12 cells washed twice with PBS, harvested and lysed. *Gaa -/-* C2C12 treated without M6P antagonist showed less variation in GAA enzymatic activity levels between different rhGAAs with statistical significance only between SSP-NewTagI-II GAA70 and Native GAA (*$p=0.05$, SEM+/-, n=3) (Figure 3.2.7 B). In contrast, *Gaa -/-* C2C12 treated with M6P antagonist and various rhGAA constructs showed statistical difference in GAA activity between SSP-NewTagI-II GAA70 and its IGF II counterpart and Native GAA (**$p<0.05$, SEM+/-, n=3), and also IGF II NewTag I construct (*$p=0.05$, SEM+/-, n=3) (Figure 3.2.7 C). This suggests that in the presence of the M6P antagonist, enzymes fitted with NewTag I or NewTag II showed improved uptake that could in part be related to uptake that bypasses the conventional M6PR pathway and therefore the use of NewTag motif as a MHP will be further interrogated.
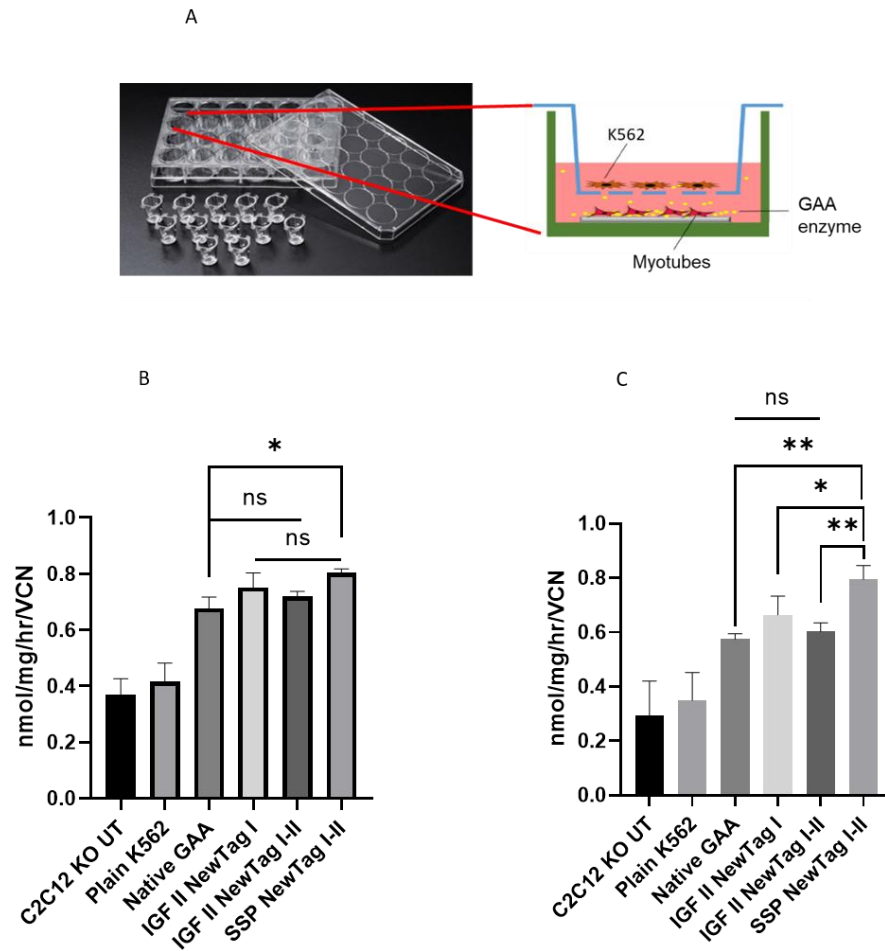
**Figure 3.2.7. Evaluation of NewTag muscle homing peptide in transwell co-culture. A)** *Transwell model used to study the uptake of different rhGAA produced by transduced K562 cells in Gaa -/- C2C12 cells. Gaa -/- C2C12 cells were seeded and differentiated on matrigel coated bottom well. $1x10^6$ K562 cells bearing different variants of rhGAA were seeded in transwell insert and left for 24 hr in co-culture at 37$^o$C and 5% $CO_2$. rhGAA uptake measured as GAA activity in Gaa -/- C2C12 in **B)** absence or **C)** presence of M6P inhibitor. C2C12 KO UT= Gaa -/- C2C12 not exposed to K562 cells; plain K562= untransduced K562 cells (*p<0.05, **p<0.01, n=3, +/-SEM, one-way ANOVA with Tukey's post hoc test). Measured GAA enzymatic activity in each sample was normalised by the sample's VCN. Normalised values were then plotted.*

### 3.2.5 Glycogen starvation assay

For the treatment of Pompe disease the entry of rhGAA into target cells as well as its lysosomal targeting are crucial in order to break down the accumulated glycogen. I showed that muscle homing peptide attached to rhGAA can facilitate the entry into *Gaa -/-* C2C12 myotubes even in presence of M6PR competitive antagonist. To determine the ability of the muscle homing peptide tagged rhGAA to cleave lysosomal glycogen upon uptake through CI-M6PR independent mechanism, I measured the glycogen level in treated *Gaa -/-* C2C12 cells in a starvation assay. *Gaa -/-* and WT C2C12 cells were seeded in transwell plates to determine a glycogen baseline level in normal physiological conditions and also under a starvation stress. Upon differentiation into myofibers, cells were starved for 24 hrs. Then, K562 cell line expressing different rhGAA constructs with similar values of VCN (3 ± 0.3) were counted and placed inside transwell inserts. C2C12 cells and K562 cells were incubated with or without 5mM M6P competitive antagonist. Co-culture was left for additional 48 hrs in glucose and pyruvate free medium. After this incubation time, C2C12 cells were harvested, washed twice with PBS and processed for glycogen detection.

Baselines of glycogen content were set up for WT and untreated *Gaa -/-* C2C12 cells in normal or starved culture conditions (Figure 3.2.8 A). When glycogen content was measured in starved *Gaa -/-* C2C12 cells exposed to different rhGAA constructs without M6P inhibitor, these cells showed significant glycogen reduction compared to control *Gaa -/-* C2C12 cells with SSPNewTagI-IIGAA construct being the most effective in reducing glycogen by 78% (Figure 3.2.8 B). However, when glycogen reduction was compared to starved *Gaa -/-* C2C12 treated with Native GAA protein the difference become less apparent and not significant for neither of IGF NewTag bearing GAA constructs. Only SSP NewTagI-IIGAA construct showed significant difference in reduction of glycogen compared to a reduction caused by Native GAA secreted protein (**$p<0.01$, +/-SEM, n=3). The SSP NewTagI-IIGAA caused about 50% lower glycogen level detected compared to the one of Native GAA treated group (Figure 3.2.8 B). Also starved *Gaa -/-* C2C12 treated with IGF NewTagI-IIGAA showed 20% lower glycogen level compared to cells treated with Native GAA (Figure 3.2.8 B).

Starved *Gaa -/-* C2C12 treated with M6P inhibitor and exposed to different rhGAA construct showed significant difference to control and plain K562 group (***p<0.001, +/-SEM, n=3) (Figure 3.2.8 C). The highest glycogen reduction was measured again in cells exposed to SSPNewTagI-II-GAA protein which well correlated to the higher level of GAA enzymatic activity observed in SSPNewTagI-IIGAA exposed *Gaa -/-* C2C12 cells. Interestingly all NewTag constructs, at this time, showed significant difference compared to Native GAA treated group, ranging between 70% for SSPNewTagI-II GAA to 50% with IGF NewTagI-IIGAA (Figure 3.2.8 C). However, this should not come as a surprise since Native GAA's cellular internalisation/uptake relies only on M6PR and as this receptor was blocked no internalisation occurred.

Furthermore, on average 20% greater glycogen reduction was seen (+/- M6P) in starved *Gaa -/-* C2C12 cells treated with NewTagI-IIGAA construct, which secretion was supported by SSP compared to IGF.

These findings suggested that NewTag peptide allows both the uptake of rhGAA through plasma membrane and its lysosomal targeting. Starved *Gaa -/-* C2C12 cells, without M6P inhibitor, treated with SSPNewTagI-IIGAA construct showed 50% greater reduction of glycogen accumulation compared to cells treated with Native GAA and over 70% reduction of glycogen in presence of M6P inhibitor. This set of evidence allowed to suggest that SSPNewTagI-IIGAA might be a potential candidate for gene therapy in PD.
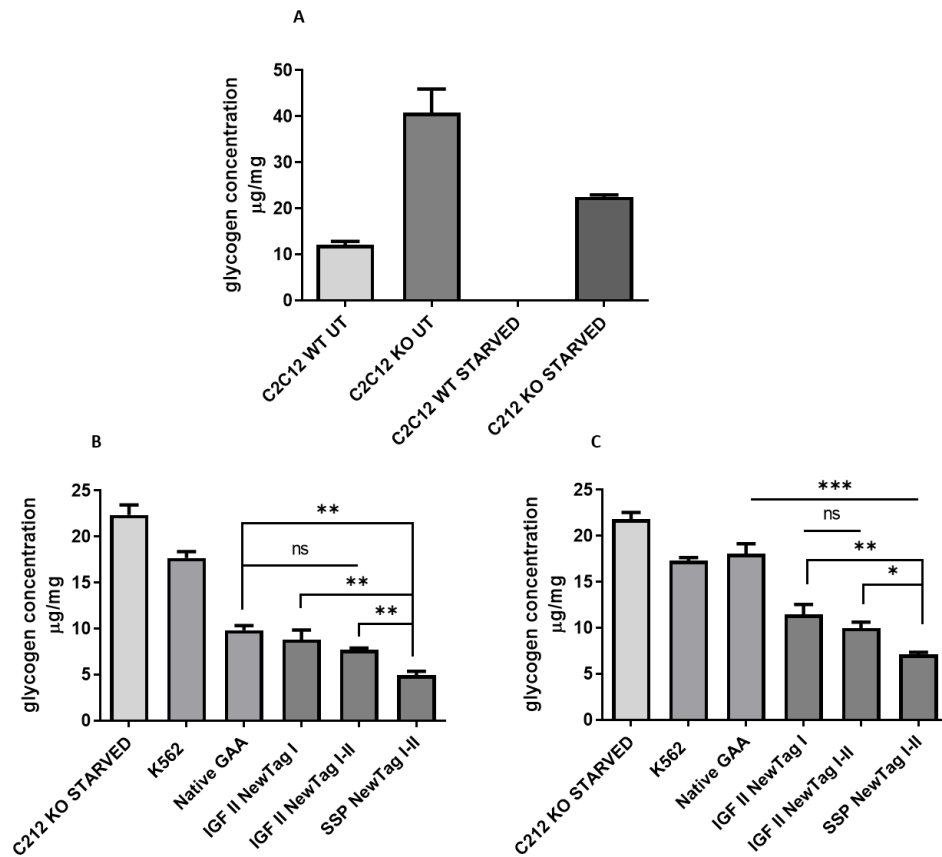
***Figure 3.2.8. Analysis of lysosomal glycogen accumulation in Gaa -/- C2C12 cells after uptake of rhGAA constructs***. *A) After 48 hrs in starvation conditions, WT and Gaa -/- C2C12 cells were harvested to define a baseline for physiological levels of glycogen accumulation (n=2, +/- SEM). **B)** In Gaa -/- C2C12 exposed to different rhGAA, without the presence of M6P inhibitor, there was significant glycogen reduction compared to native GAA seen in sample exposed to SSPNewTagI-II construct, reaching 78% reduction in glycogen compared to starved untreated cells and 50% more reduction in glycogen compared to Native GAA treated group. **C)** In Gaa -/- C2C12 cells exposed to different forms of rhGAA with M6P inhibitor, a range of glycogen reduction level was observed, with Gaa -/- C2C12 cells exposed to SSPNewTagI-IIGAA showing over 70% reduction in glycogen compared to untreated starved Gaa -/- C2C12 cells and Native GAA treated group, followed by 54% reduction in glycogen in IGFII NewTagI II GAA exposed cells ($***p<0.001$, $**p<0.01$, $*<0.05$, +/-SEM, n=3, one-way ANOVA with Tukey's post hoc test). C2C12 WT/KO UT = WT/KO C2C12 in full medium without exposure to K562 cells,*

*C2C12 WT/KO STARVED = WT/KO C2C12 in reduced (starvation) medium without exposure to K562 cells. K562 = untransduced K562 cells.*

### 3.2.6 Discussion chapter

Under normal physiological conditions about 10% of GAA protein intracellular production escapes through GI apparatus into extracellular space. In gene therapy modified cells, GAA protein, under the drive of the SSP signal peptide, a substantial majority of rhGAA was exported extracellularly. Posttranslational modification i.e., glycosylation of our GT rhGAA closely resembled that seen of native GAA protein. This was greatly reassuring that N-terminus implemented modification did not impact posttranslational N-linked oligosaccharides composition.

In addition, transgene promoters PGK vs LCR-EFS1 were evaluated for the best transcriptional GAA transgene performance. In addition, transgene promoters PGK vs LCR-EFS1 were evaluated for the best transcriptional GAA transgene performance. Out of two tested transgene promoters the LCR-EFS1 tandem showed higher overall rhGAA production and secretion in K562 cell line. This cellular model is widely used *in vitro* for erythroid specific gene assessment due to the globin gene expression (Rutherford et al., 1979; Uchida et al., 2018) and also it is very desired model to test LCR part of our transgene promoter. However, *in vitro* models sometimes do not reflect a genuine *in vivo* gene expression (Calabria & Shusta, 2008). Therefore, in order to understand a potential *in vivo GAA* transgene expression levels, under PGK vs LCR-EFS1 promoter, transduced CD34[+] and/or murine lineage negative cells and further differentiated into myeloid lineages and/or erythroid progeny could be used and studied. However due to restricted timeline of this project optimisation and generation of transduced hematopoietic progeny cells *in vitro* was unfeasible.

CRISPR Cas9 generated C2C12 *Gaa-/-* myotubes murine cell line allowed further investigation of NewTagI-IIGAA protein uptake and lysosomal targeting. I initially established a transwell system that was able to deliver rhGAA cross-correction to C2C12

*Gaa-/-*. Furthermore, utilizing a starvation protocol (Chapter 3.2.5) in the transwell setting, showed glycogen reduction in starved C2C12 *Gaa-/-* treated with several different form of NewTag GAA recombinant proteins.  However, I observed greater rhGAA activity and reduction in glycogen in C2C12 *Gaa-/-* treated with SSPPNewTagI-IIGAA construct. However, the *in vitro* cross-correction observed was of small amplitude, without full glycogen reduction to WT level. At this point it was difficult to determine if this level of glycogen decrease observed *in vitro* would be sufficient to translate into full correction of PD in a *vivo* mouse model.

However, given the correct rhGAA posttranslational modification followed by high level of secretion of the protein and ability of NewTagI-II domain to exercise rhGAA internalisation and lysosomal targeting in treated C2C12 *Gaa-/-*, it was plausible to continue this study to an *in vivo* evaluation using SSP NewTagI-II GAA construct.

## 3.3 *In vivo* assessment of GAA activity and glycogen burden in treated Pompe mouse model.

Previous HSCs gene therapy studies on Pompe disease (Piras et al., 2020; Stok et al., 2020; van Til et al., 2010) have led to a conclusion that bone marrow transplant, with lentiviral vector transduced HSCs, indeed reduces lysosomal glycogen burden. These studies also showed that cross correction is a valuable mechanism to provide cells with missing/dysfunctional enzymes and correct the disease. However, van Til *et al.* and Stok *et al.* used SFFV promoter, derived from long terminal repeats (LTRs) of Spleen Focus Forming Virus, in order to achieve high level of transgene expression and in consequence provide therapeutic amount of GAA protein; however, this promoter has been shown to have transactivating potential in IVIM (*in vitro* immortalisation) assays does not guarantee a clinical safety. Viral promoters, due to the presence of transactivating enhancer sequences, are known to upregulate any oncogenes being in close proximity (Milone & O'Doherty, 2018). Furthermore, Stok *et al.* (Stok et al., 2020) showed that VCN of 7 in transduced HSCs was able to deliver therapeutic level of GAA enzyme in blood plasma and provide correction of Pompe disease in experimental animals.

We hypothesised that enhanced GAA transgene expression under the transcriptional control of the βLCR EFS1α promoter would drive sufficient gene transcription to deliver therapeutic correction of Pompe disease. In addition, promoters used in my study being mammalian promoters without strong transactivating potential would lead to a safer clinical safety profile. Furthermore, optimised N-terminus rhGAA with MHP and powerful SSP was hypothesised to provide sufficient therapeutic enzyme level and correct lysosomal glycogen storage *in vivo*.

### 3.3.1 HSCs purification, transduction, and transplantation to *Gaa -/-* recipient mice.

Experimental steps of HSCs gene therapy process are depicted in Figure 3.1.1.
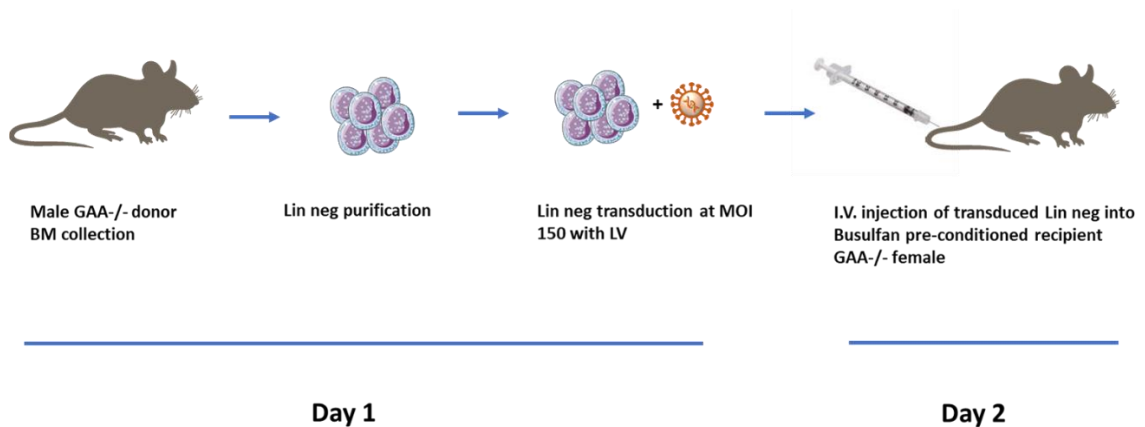
**Figure 3.3.1.** *Schemata of collection and processing HSCs. Day 1 includes harvesting male Gaa -/- donor BM (bone marrow), followed by purification to obtain pure population of lineage negative (lin neg) cells, lentiviral transduction at MOI 150. On day 2 transplantation of transduced Lin neg cells into Busulfan pre-conditioned female Gaa -/- recipient mice take place.*

The experimental layout of *in vivo* trial is as follow:

➢ Recipient mice receive dose of 25mg/kg/day busulfan IP injection for 4 consecutive days

➢ 24 hour resting window between last injection of busulfan and transplant of transduced HSCs.

➢ 24 hr rest period follows infusion of the transduced HSCs into recipient mice.

$1x10^5$ transduced HSCs is kept *in vitro* for further analysis i.e., liquid culture (LC) and progeny development (CFU) (M&M section 2.12.2 and 2.12.3 respectively). Liquid culture is analysed after 7 post-transduction and is used for analysis of VCN and GAA activity, followed by day 14 with the assessment of differentiation of individual progenitor cells into colonies, and analysis of VCN and GAA activity in pooled colonies.

BM of *Gaa -/-* male donor mice were harvested and purified to obtain a selected population of HSC lineage negative cells, positive for c-kit (CD117) and Sca-1 markers of

stemness. Lineage negative (Lin neg) represents population of naïve HSCs capable to give rise to all cellular blood lineages when transplanted into recipient animal (Goldberg et al., 2014; Sawen et al., 2018). Further FACS analysis of lin neg purity was conducted showing good levels of purity reaching 81.54% pure lin neg population (Figure 53.3.2 D) with different expression of stemness markers  c-Kit and Sca-1 (Figure 3.3.2 E).
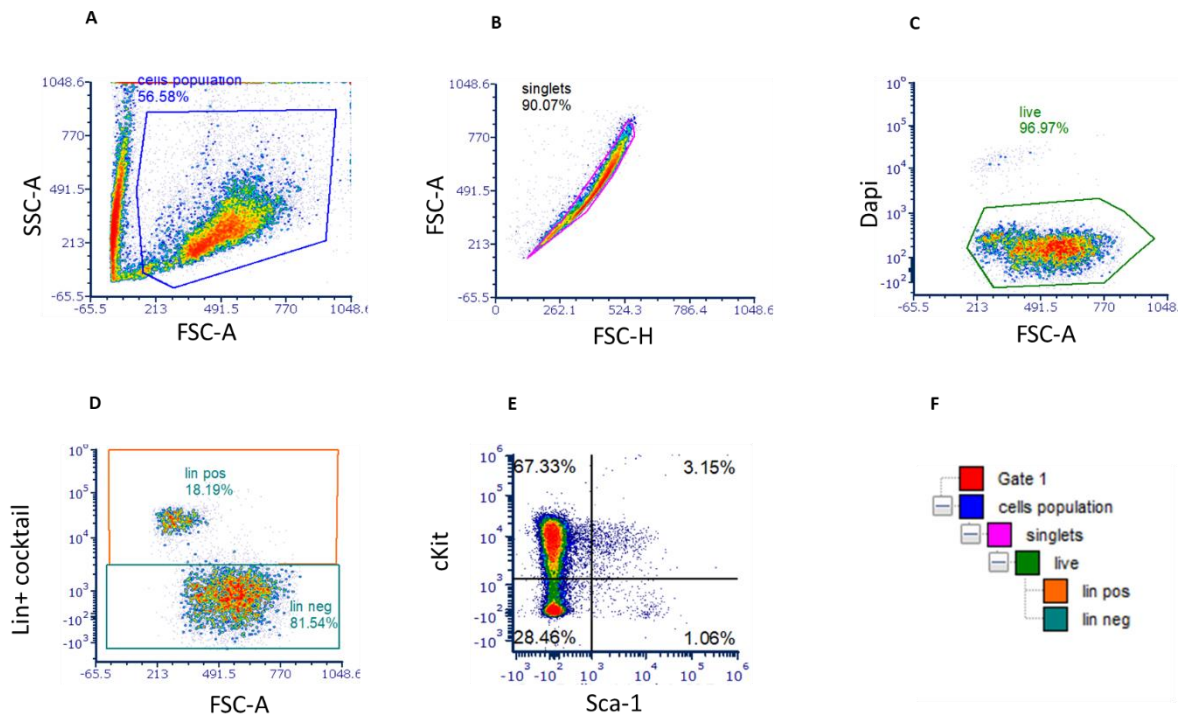


*Figure 3.3.2. Representative Lin neg purity and gating strategy. Assessment of purity of Lin neg elution fraction was conducted by FACS gating for **A)** side scatter (SSC-A) vs forward scatter (FSC-A), **B)** single cell population, **C)** Dapi live vs dead cells and **D)** to cocktail antibodies toward lineage positive (Lin+) specific markers stained by anti-Streptavidin antibody was used to check purity of lin neg fraction showing that in eluted fraction of lin neg cells there is 81.54% negative cells and 18.19% impurity of lin positive. **E)** cKit (PE) and Sca-1 (Violet 605) stemness markers were used to determine composition of Lin neg population from gate D (lin neg gating). **F)** Gating hierarchy shows dependence of gating strategy.*

Lin neg cells were transduced with Lentiviral vectors carrying LCR-EFS-SSP-NewTagI-II-GAA(Δ70) and also LCR-EFS1α-GAA, MOI of 150 as described in materials and methods section 2.10.1. Then cells were harvested the following day, washed and transplanted into female conditioned *Gaa -/-* recipient mice at ~ $6x10^5$ cells per animal.

In addition, a liquid culture (LC) with $1x10^5$ cells and colony forming unit culture (CFU-c) with $3.1x10^3$ transduced HSCs were seeded. Both LC and CFU-c were set up to determine GAA activity along with VCN in transplanted material. CFU-c in addition to enzymatic activity and VCN was also used to quantitate and characterize hematopoietic progenitors from transduced and transplanted HSCs.

LC collected after a week in culture shown VCN of 1.12 for SSPNewTagI-II and 1.15 for GAA native, and GAA activity of 38 and 114.51 nmol/mg/hr, respectively (Table 3.3.1 A). CFU-c were collected two weeks post transduction showing VCN of 1.35 for SSPNewTagI-II and 1.45 for GAA native, and GAA activity of 35.66 and 234.69 nmol/mg/hr, respectively (Table 3.3.1A). Irrespective of similar values of VCN, GAA activity recorded in cells transduced with GAA native construct showed greater value in contrast to SSPNewTagI-II. This phenomenon can be explained since GAA native proteins do not possess a signal peptide and therefore accumulates in transduced cells in contrast to proteins equipped with SSP signal peptide.

Transplanted *Gaa -/-* animals with transduced cells after 3 months showed no significant higher level of rhGAA protein in blood plasma compared to *Gaa -/-* untreated mice. There was no significant reduction in glycogen concentration in collected tissues and there was no significant difference in skeletal muscle performance in functional rota rod test between gene therapy treated and untreated *Gaa -/-*. In addition, collected BM samples showed very low VCN ~1, which would explain poor GAA correction and activity.

Despite the fact that LVs are capable of integration of up to 9kB transgene (Fassler, 2004), the total size of transgene and transgene promoter influence viral performance leading to lower VCN in transduced cells (Sweeney & Vink, 2021; Terwilliger et al., 1989).

In our study the core of the GAA transgene fragment could not be modified/shortened as its functionality would be affected. However, a possible modification and reduction in the total size of transgene cassette and its promoter was still able to be achieved by replacing LCR-EFS 2.5kB for PGK 500bp transgene promoter. It was hoped that reduction in genetic material carried by used vector will enhance viral machinery assembly and integration of transgene into genomic material of a host.  We hypothesised that a higher number of copies of the transgene per transduced cell would generate, under constitutive PGK promoter, more mRNAs followed by more functional copies of rhGAA protein, compared to lower number of copies under tandem of LCR-EFS. Our hypothesis was to be tested *in vivo* with level of rhGAA activity detected in plasma of transplanted *Gaa-/-* animals.

LV-PGK-SSP-NewTagI-II GAA and LV-PGK-GAA native were designed (Figure 3.3.3), cloned and produced.
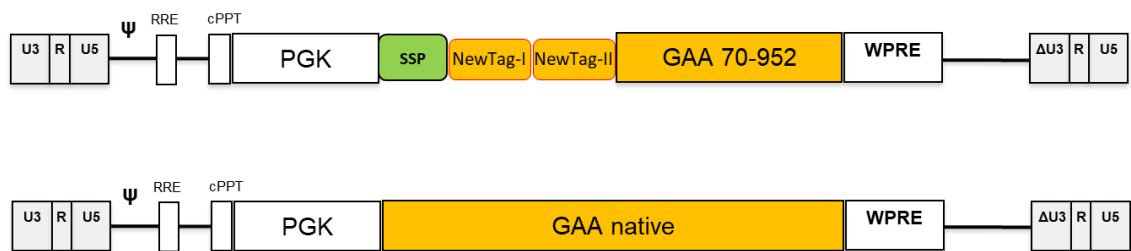


***Figure 3.3.3. Substitution of transgene promoter.*** *Structural organisation of lentiviral vectors with top presenting PGK promoter followed by SSP-NewTagI-IIhGAA transgene, and bottom showing PGK promoter followed by native hGAA gene only. Both transgenes and PGK promoter were cloned in pCCL lentiviral vector.*

HSCs cells were harvested from donor *Gaa -/-* mice and purified. For both constructs, MOI of 150 was used to transduced harvested HSCs. Cells were left with lentiviral particles overnight in transfection medium (as previously described). LC and CFU-s were set up as previously.

A week post transduction data from LC and two weeks for CFU-c showed a significant increase in VCN with both LV-PGK constructs (Table 3.3.1 B). In addition, CFU-c did not show any major differences in hematopoietic progenitors' colonies from transduced HSCs compared to un-transduced HSCs (Figure 3.3.4 B). This data led to a conclusion that further study should be continued with LV-PGK instead LV-LCR-EFS constructs as this would allow to obtain higher VCN in transduced HSCs and lead to potentially better biochemical and therapeutic outcomes.

Table 3.3.1. LC[1] and CFU-c [2] early VCN and GAA activity in transduced HSCs with **A)** LCREFS1α and **B)** PGK promoter.

A

| Sex | LV construct | Lin-ve LC[1] | | Lin-ve CFU-c[2] | |
|---|---|---|---|---|---|
| | | VCN | GAA Activity (nmol/mg/hr) | VCN | GAA Activity (nmol/mg/hr) |
| F | LCR EFS SSP NewTagI-II GAA | 1.12 | 38 | 1.35 | 35.7 |
| | LCR EFS GAA Native | 1.15 | 114.5 | 1.47 | 234.7 |
| | UT KO | 0.02 | 1.3 | 0.03 | 2.72 |

B

| Sex | LV construct | Lin-ve LC | | Lin-ve CFU-c | |
|---|---|---|---|---|---|
| | | VCN | GAA Activity (nmol/mg/hr) | VCN | GAA Activity (nmol/mg/hr) |
| F | PGK SSP NewTagI-II GAA | 5.95 | 360 | 8.63 | 915.9 |
| | PGK GAA Native | 4.64 | 716.7 | 4.44 | 1433.3 |
| | UT KO | 0.07 | 1.9 | 0.06 | 3.43 |

---

[1] LC – liquid culture

[2] CFU-c – colony forming unit culture
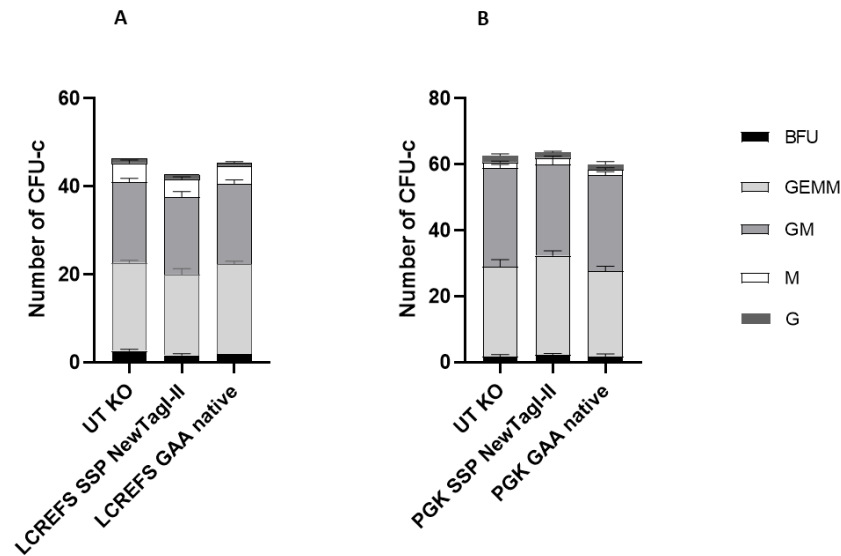
***Figure 3.3.4. Colony forming unit culture's analysis.*** *CFU-c counts for untreated and lin neg cells transduced with **A)** LCREFS1α and **B)** PGK bearing constructs. Three 3cm plates each containing 1.1x10^3 cells were seeded in 1ml of Methocult and allowed to grow and differentiate for two weeks in 37°C and 5% $CO_2$. Each plate was inspected, and colonies formed counted including: **BFU-E** representing primitive erythroid progenitor cells, **GEMM** common myeloid progenitors, **GM** known as granulocyte–macrophage progenitors, **M** macrophage and **G** granulocyte monopotent progenitors. HSCs transduced with any of these constructs showed no significant difference in numbers of colonies counted compared to correspondent untreated Gaa -/- HSCs (n=3, +/- SEM). UT KO = untreated Gaa -/- HSCs.*

Transplant of purified and transduced HSCs from *Gaa -/-* mice (as previously described) with both LVs bearing PGK promoter with MOI of 150 was conducted. This experiment was set up to determine if reducing the size of LV would boost VCN in transduced HSCs but also on the other hand the new promoter would maintain the need of high transgene expression in vivo.

Prior to transplant with transduced HSCs, recipient mice went under Busulfan regime of 25mg/kg/day (as per protocol in section materials and methods 2.10.5) to myeloablate recipients' BM and allow engraftment of the transduced and transplanted HSCs. Transplanted animals were allowed to recover on wet diet for two weeks following 3 months on solid diet. After this period experiments were terminated with final collection, from each transplanted animal, of blood, BM, and several types of tissues for further investigation of the GAA activity and glycogen content.

### 3.3.2 Biochemical and functional analysis

To assess biochemical and functional correction in *Gaa -/-* mouse model, 5 animals received PGK-SSP-NewTagI-II-GAA transduced HSCs and 4 animals received PGK-GAA transduced HSCs. In addition, 8 wild type and 6 GAA knockout untreated animals were used as control groups. All animals used in this study were 6 - 8 weeks of age at the beginning of the *in vivo* experiment. The following experimental plan was set to interrogate correction of Pompe biochemical and functional phenotype in gene therapy *Gaa -/-* mice:

1. **Engraftment and VCN**
   a. BM
   b. PBMC
2. **Cellular blood lineages development**
3. **GAA activity level**
   a. BM
   b. PBMC
   c. Blood plasma
   d. Harvested tissues
4. **Pompe disease hallmarks levels**
   a. Glycogen in harvested tissues
   b. Glucose tetrasaccharide (Glc4) in collected urine
5. **Post-translational glycosylation modification of rhGAA purified from blood plasma**
6. **Functional test**

a. Muscle coordination tested by rota rod

b. Correction in heart muscle measured by echocardiography


## 3.3.2.1 Engraftment and VCN.

A level of engraftment was calculated by using qPCR probes against Y chromosome as a target gene and Titin as a reference gene. qPCR reaction was run using purified DNA samples from PBMC and BM for each individual animal. The values of engraftment presented on Figure 3.3.5 A show 100% of engraftment in gene therapy groups and 0% Y chromosome reading in WT and *Gaa -/-* mice as expected.

In addition, VCN was assessed using the same samples of purified DNA from both BM and PBMC for each experimental animal and shown in Table 3.3.2 and Figure 3.3.5 B and C, respectively. Final termination VCN data is consistent with VCN shown at LC and CFU-c of the transplanted HSCs indicating a stable expression of the transgene in transduced HSCs over time.


**Table 3.3.2. VCN detected in BM and PBMC cells from gene therapy and control groups.**

| | Animal ID | UT KO | UT WT | PGK SSP NewTagI-IIGAA | PGK GAA Native |
|---|---|---|---|---|---|
| BM | 1 | 0 | 0 | 9.6 | 4.2 |
| | 2 | 0 | 0 | 7.2 | 4.9 |
| | 3 | 0.1 | 0.1 | 6.5 | 4.7 |
| | 4 | 0 | 0.1 | 8.2 | 6.4 |
| | 5 | 0.1 | 0.1 | 5.9 | |
| | 6 | 0 | 0.1 | | |
| | 7 | | 0 | | |
| | 8 | | 0.1 | | |
| PBMC | 1 | 0.1 | 0 | 8.5 | 4.5 |
| | 2 | 0 | 0 | 5.4 | 4.2 |
| | 3 | 0 | 0.1 | 8.9 | 3.6 |
| | 4 | 0.1 | 0 | 8.4 | 5.7 |
| | 5 | 0 | 0.1 | 4.2 | |
| | 6 | 0.1 | 0.1 | | |
| | 7 | | 0 | | |
| | 8 | | 0 | | |
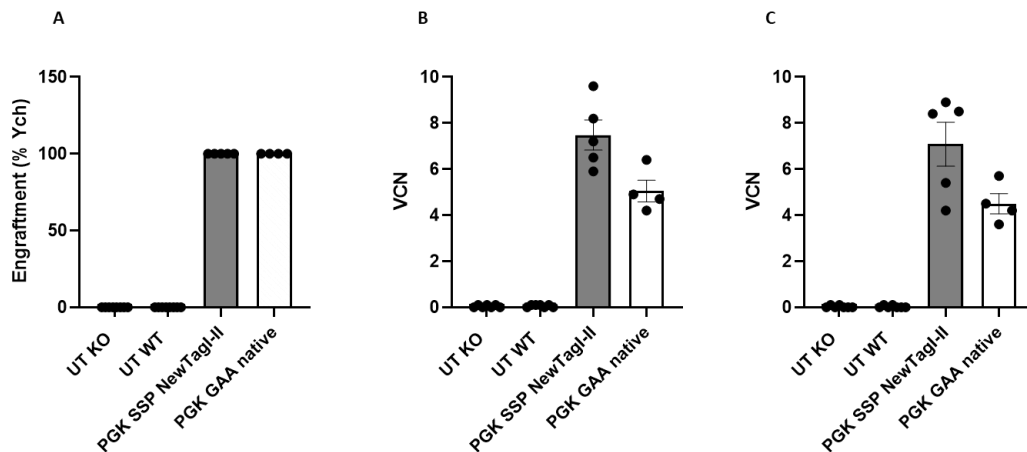
***Figure 3.3.5. End point engraftment (%) and VCN analysis.*** *Gene therapy recipient female mice were transplanted with donor males HSCs. Genomic DNA from harvested bone marrow per each experimental animal was analysed by using qPCR assays toward Y chromosome carried only by males' HSCs transplanted into female recipients.* ***A)*** *UT KO and UT WT showed no signal identified by qPCR Y chromosome assay authenticating this method of detection of engraftment. In contrast, both gene therapy groups showed 100% Y chromosome normalised by housekeeping gene Titin. In addition, VCN* ***B)*** *in BM and* ***C)*** *in purified PBMC from peripheral blood of both gene therapy groups showed similar levels of VCN seen at earlier timepoints LC and CFU-c stages confirming persistence of transgene expression and hence lack of viral vector silencing in gene therapy animals. (UT KO n=6, UT WT n=8, PGK SSP NewTagI-II n=5 and PGK GAA native n=4 animals, +/-SEM)*

## 3.3.2.2 Cellular blood lineages

Further analysis of cellular blood lineages was undertaken to determine if expression of GAA from any of the used LV vectors caused any disturbance to lineage development compared to untreated groups. For this purpose, fluorescence-activated cell sorting (FACS) analysis was used. Utilising cellular surface markers (B220 - B lymphocytes, CD3 – T lymphocytes, CD11b and Gr1 – macrophages, however Gr1 staining was unsuccessful, and no data was collected for this marker) carried on harvested bone marrow cells, showed no significant differences in distribution of blood lineages

between gene therapy treated groups and control animals (Figure 3.3.6). This data indicated that viral vectors used to transfer transgene did not affect lineages population development formed from transplanted HSCs compared to untransduced control BM from *Gaa -/-* mice.
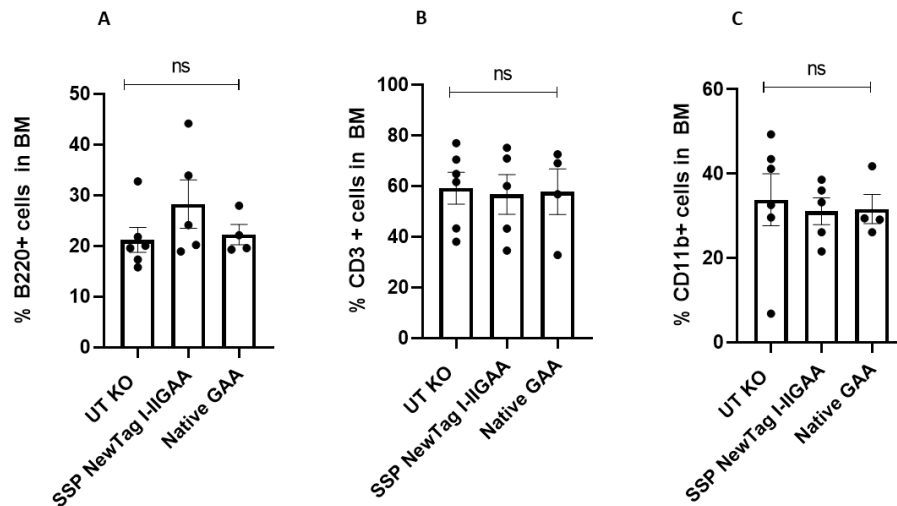


***Figure 3.3.6. Differentiation of transplanted HSCs into blood lineages.*** *Bone marrow cells were harvested from experimental animals and 1x10^5 cells stained for several cellular markers including **A)** B220, **B)** CD3 and **C)** CD11b to indicate differentiation profile of transplanted HSCs. Harvested bone marrow from UT KO, GAA native and SSPNewTagI-II GAA showed similar percentages across all measured cell populations, no statistical difference was observed (+/-SEM, UT KO n=6, SSP NewTagI-II GAA n=5 and GAA native n=4,* Kruskal-Wallis H test*).*

### 3.3.2.3 GAA level in PBMC, BM, blood plasma and harvested tissues

GAA activity was measured using 4MU substrate and the values were presented as nmol/mg/hr for PBMC, BM and harvested tissues and as nmol/ml/hr for GAA measured in blood plasma.

Harvested and lysed BM and PBMC shown high level of GAA enzymatic activity in both gene therapy groups with moderate decreased in SSP-NewTagI-IIGAA group (Figure 3.3.7 A and B respectively); this was expected and confirmed *in vitro* findings in K562 transduced cell line. Enzymatic activity of GAA in plasma of SSPNewTagI-IIGAA cohort animals showed superior higher level compared to all other groups (***p<0.001, +/-SEM) (Figure 3.3.7 A). In addition, both transgenes generated higher level of GAA enzymatic activity in BM compared to UT KO and UT WT (***p<0.001, +/-SEM) (Figure 5.8 A). Also, PBMC showed statistically higher GAA enzymatic activity for SSPNewTagI-IIGAA, and GAA Native compared to UT KO and UT WT groups (***p<0.001, +/-SEM) (Figure 3.3.7 B).

The last but not least was the determination of GAA enzymatic activity in harvested and homogenised tissues of experimental animals. These samples covered the most affected tissues in Pompe disease including heart, diaphragm, gastrocnemius-soleus (GAS), tibialis anterior (TA), brain but also, we tested lungs and liver.

Figure 3.3.7 C and D demonstrated various levels of GAA activity measured in homogenised tissues coming from experimental animals WT, *Gaa -/-*, SSP-NewTagI-IIGAA, and GAA native. Majority of harvested and homogenised tissues and in particular muscular tissues show low level of GAA activity in gene therapy treated groups compared to UT WT level of enzymatic activity (Figure 3.3.7 C). Both PGK-SSPNewTagI-IIGAA and GAA native show statistical significance in GAA activity level compared to KO UT homogenised samples, particular in muscular tissues i.e., heart, diaphragm, tibialis anterior, and gastrocnemius (*p<0.05, n=5 for PGK SSP NewTagI-IIGAA and n=4 for GAA native, +/-SEM). Also, brain tissues showed statistical difference in GAA activity between both constructs and UT KO (*p<0.05, n=5 for PGK SSP NewTagI-IIGAA and n=4 for GAA native, +/-SEM) however the levels achieved by the gene therapy groups are much lower than levels seen in UT WT animals (Figure 3.3.7 D).
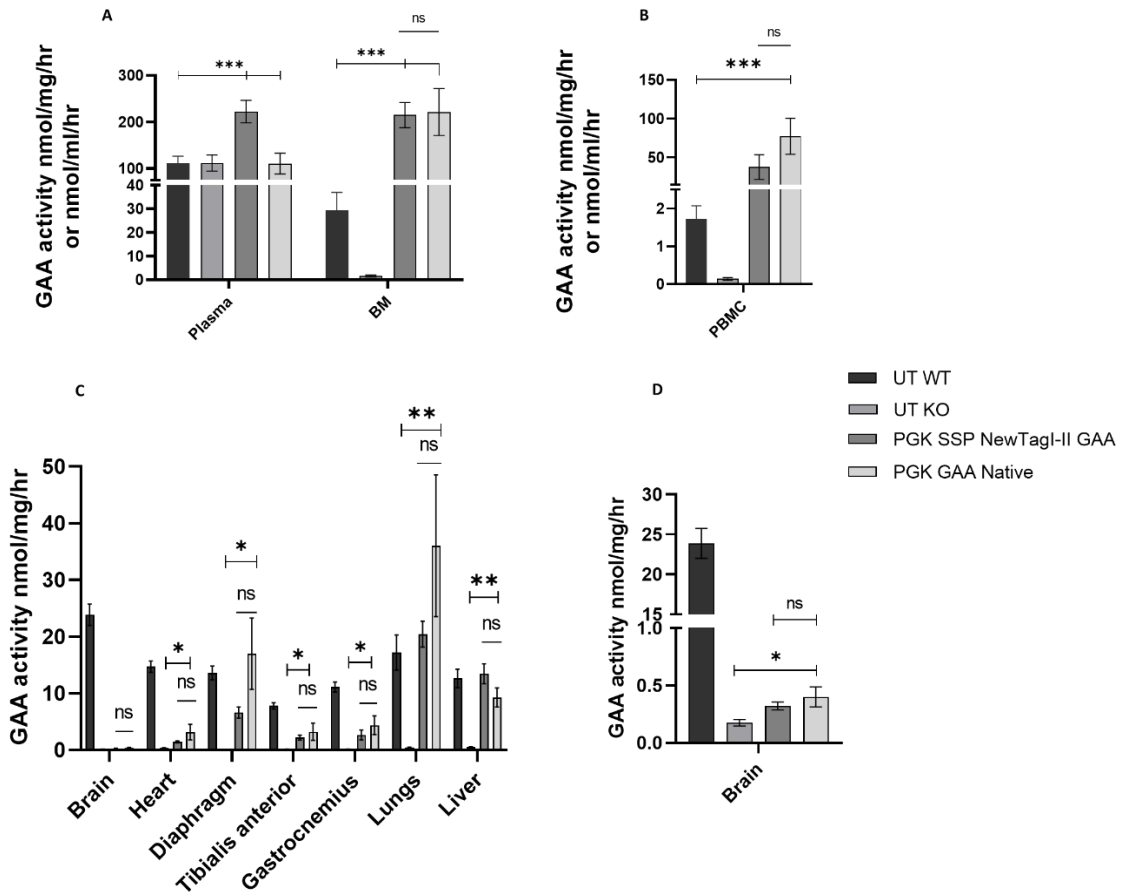
***Figure 3.3.7. GAA activity in harvested blood, plasma and tissues. A)*** *Enzymatic activity presented in blood plasma showed superior higher level in animals transplanted with PGK-SSPNewTagI-IIGAA construct compared with remaining experimental arms (\*\*\*p<0.001, +/-SEM) and this observation was in accordance with in vitro experimental data generated from culture medium of K562 cells transduced with SSPNewTagI-IIGAA construct. GAA activity in BM showed significantly higher level compared to KO UT for both constructs (\*\*\*p<0.001, +/-SEM). GAA activity in plasma expressed as nmol/ml/hr and BM as nmol/mg/hr.* ***B)*** *Both gene therapy groups generated statistically higher GAA activity in PBMC compared to WT UT and KO UT (\*\*p<0.001, +/-SEM).* ***C)*** *GAA activity recorded from homogenised tissues. Gene therapy animals' samples showed statistically higher GAA activity in heart, diaphragm, tibialis anterior and gastrocnemius (\*p<0.05, +/-SEM) and lungs and liver (\*\*p<0.01, +/-SEM) compared to UT KO group.* ***D)*** *GAA activity detected in homogenised brain showed significant higher enzymatic activity in gene therapy animals compared to untreated Gaa -/- (\*p<0.05, +/-SEM). n=5 for PGK SSP*

*NewTagI-IIGAA, n=4 for GAA native, n=6 UT KO, and n=8 for UT WT. Groups were analysed using one-way ANOVA with Tukey's post hoc test.*

**3.3.2.4 PD hallmarks - glycogen and glucose tetrasaccharide (Glc4) level**

Following GAA enzymatic activity, a glycogen analyses took place on homogenised tissues. The analysis of glycogen in each individual tissue of every experimental animal revealed the extent to which the gene therapy and N-terminus modification allowed to reduce glycogen accumulation compared to untreated *Gaa -/-* animals.

Both gene therapy groups, SSP-NewTagI-IIGAA group and GAA native, showed significant reduction in glycogen level compared to *Gaa -/-* cohort including in brain, heart, diaphragm, GAS and TA (Figure 3.3.8 A, B). This reduction ranged roughly from 50% seen in diaphragm to 35% of UT levels in heart. Muscular tissues i.e., Tibialis anterior and gastrocnemius showed reduction in glycogen reaching 40-50 % compared with UT KO. However, no significant differences in glycogen reduction were observed between SSP-NewTagI-IIGAA and GAA native treated groups. Furthermore, none of the gene therapy treated groups reduced glycogen level to that of seen in WT cohort.

GAA activity seen is some GAA native treated animals in diaphragm and lungs did not translate into lower glycogen readings in corresponding tissues. We hypostatise that perhaps insufficient organ's flushing with PBS took place leaving some rhGAA circulating in blood still present in homogenised organs contributing to a total detected GAA activity.
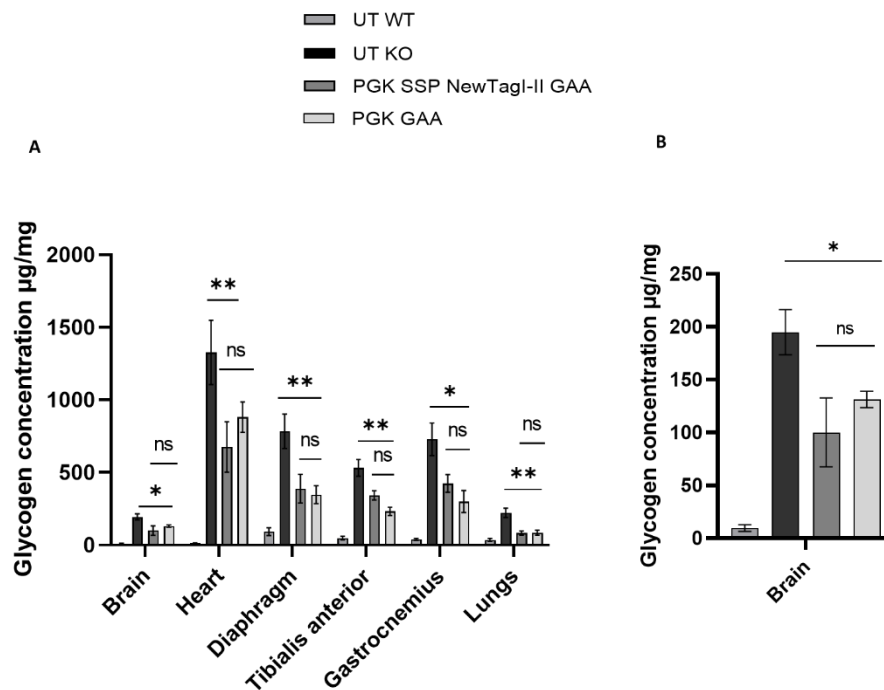
**Figure 3.3.8. Glycogen concentration in homogenised tissues. A)** *Both gene therapy groups showed significant reduction in detected glycogen in harvested tissues compared to UT KO animals (\*p<0.05, \*\*p<0.01, +/-SEM).* **B)** *Also, brain samples showed significant reduction in glycogen concentration between gene therapy treated animals and UT KO (\*p<0.05, +/-SEM) with SSPNewTagI-IIGAA showing a positive trend to lower glycogen readouts compared to GAA native. However, there was no significant difference seen in between PGK-SSPNewTagI-IIGAA and GAA native treated animals in any of the tissue category. (n=5 for PGK SSP NewTagI-IIGAA, n=4 for GAA native, n=6 UT KO, and n=8 for UT WT, one-way ANOVA with Tukey's post hoc test was used to analyse experimental groups).*

In addition, glucose tetrasaccharide (Glc4) level in mice's urine was measured and data provided by our collaborators at Great Ormond Street Hospital NHS Laboratory Medicine Department.

Glc4 is a product of degradation of glycogen and it is one of the biomarkers of Pompe disease (Kishnani, Steiner, et al., 2006; Tarnopolsky et al., 2016). In Pompe patients'

urine GlC4 level is very high reflecting elevated level of glycogen stored (Kishnani, Steiner, et al., 2006). *In vitro* analysis of Glc4 concentration in UT KO animals showed high level of this biomarker when compared to relatively low level seen in UT WT animals. Gene therapy animals fall in between UT KO and UT WT with a level of Glc4 in SSPNewTagI-II GAA treated animals showing nearly significantly lower level at p=0.08 (+/-SEM, n=5) and GAA native animals showing no significant level of difference compared to UT KO (Figure 3.3.9). Interestingly Glc4 level measured in animals' urine showed significantly lower values in SSPNewTagI-II GAA treated animals compared to GAA native group (*p<0.05, +/-SEM, n=5 SSPNewTagI-II GAA and n=4 GAA native) (Figure 3.3.9).
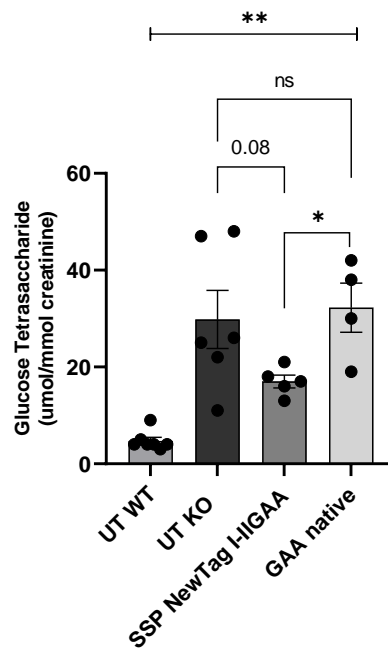


**Figure 3.3.9. *Glucose Tetrasaccharide (Glc4) level in urine.*** *3 months post-transplant 50 µl of urine from each individual animal in gene therapy treated groups and UT groups were collected and processed to determine Glc4 level. The analysis indicated that UT KO animals showed as expected high level of Glc4 and UT WT cohort showed consistently significantly lower level to all groups (**p<0.01, +/-SEM). Urine from PGK-SSPNewTagI-II GAA treated animals showed nearly significant difference to UT KO at value of p=0.08 (+/-SEM) and statistical difference to its competitor construct GAA native (*p<0.05, +/-SEM). Glc4 level detected in animals treated with GAA native showed no significant*

*difference from UT KO animals. (n=5 for PGK SSP NewTagI-IIGAA, n=4 for GAA native, n=6 UT KO, and n=8 for UT WT,* Kruskal-Wallis H test was used to analyse groups*).*

### 3.3.2.5 Post-translational glycosylation modification of rhGAA purified from blood plasma

*In vivo* data generated from GAA activity and glycogen level in treated groups and in particular SSPNewTagI-IIGAA cohort turned to be less efficacious than that predicted by the *in vitro* study. We hypothesised that posttranslational modification and in particular glycosylation might be affected by any component of modified N-terminus of rhGAA. To verify our hypothesis, I decided to study the N-linked glycoprotein patterns of rhGAA from experimental animals' blood plasma. Using Concanavalin A (Con A) agarose resins the animals' plasma glycoproteins were purified and concentrated using Amicon® Ultra 0.5 mL Centrifugal Filters. Purified and concentrated glycoproteins were then treated with glycosidases recognizing N-linked oligosaccharides, PNGase and EndoH (already described is section 3.2.1) and processed by WB with anti-human GAA antibody. To establish original molecular size of rhGAA, uncleaved samples of glycoproteins containing GAA native and SSPNewTagI-IIGAA were run alongside with glycosidases treated samples.

WB analysis with anti-hGAA antibody unveiled that both GAA native and SSP-NewTagI-IIGAA constructs presented an identical pattern of bands and band sizes in both uncleaved and cleaved/treated samples with Endo H or PNGase F (Figure 3.3.10). We hypothesised based on these findings that any of the N-terminus modification of SSPNewTagI-IIGAA peptide did not alter the N-linked oligosaccharide pattern in rhGAA and therefore the glycosylation of rhGAA protein is unlikely responsible for poor *in vivo* performance of the construct in question.
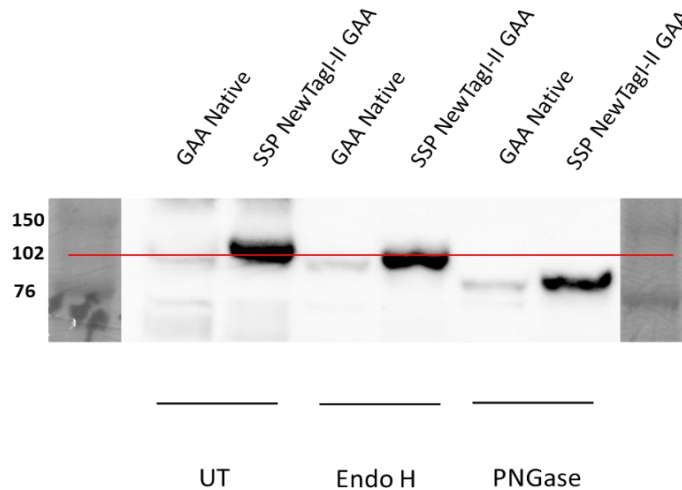
**Figure 3.3.10.** *Determination of N-linked glycosylation in rhGAA purified from blood plasma. Purified rhGAA, from blood plasma sample, from randomly chosen one representative of GAA native group and one of SSPNewTagI-II GAA animals were subject to enzymatic activity by Endo H and PNGase. Glycoproteins were loaded in each line of Western blot gel at different amount for each enzymatic condition permitted. Immunoblotting with anti-GAA showed no difference in kDa size between different fractions (UT- untreated rhGAA, Endo H and PNGase treated glycoproteins) run on Western blot membrane between GAA native and SSPNewTagI-II GAA. Red line was placed at 102 kDa of rainbow molecular marker.*

### 3.3.2.6 Functional test and Echocardiography

To assess the correction of Pompe disease in GT treated mice and evaluate muscle strength forelimb grip test and hang wire test were used. However due to a malfunction in grip test device no further assessment was able to be conducted and it was abandoned. The hanging wire test after initial trial was also halted due to inability of untreated wild type animals to grip/hang. After several unsuccessful trials collecting data from untreated WT and untreated *Gaa-/-*, with *Gaa-/-* mice griping/hanging for a longer time compared to WT (both groups were sex and age matched) it appeared impossible to generate baselines to assess a muscle correction in GT *Gaa-/-* treated groups.

An alternative analysis - the rotarod test was available, which primary can determine coordination, balance, and condition of muscle, but also muscle endurance, which allows to draw a conclusion of muscle state and its strength (Chapillon et al., 1998). Using rotarod permits to determine a functional correction in GT treated *Gaa-/-* animals vs *Gaa-/-* untreated groups. Previously troubled *Gaa-/-* untreated and WT groups responded well to rotarod, providing with meaningful values permitting to assess muscular endurance and strength in GT treated *Gaa-/-* animals.

Rotarod assembly and processing - animals from all groups, treated and controls, were placed on a wheel. This wheel accelerated gradually over time forcing animals to increase their speed, muscle workload and coordinate their position. This test is a measurement of motor coordination, balance, and muscle strength.

Each animal was allowed to adjust to the rotarod for 5 minutes in a slow, unaccelerated pace prior to 5 accelerated runs with 3 minutes break between each turn, however two initial runs for each animal were discarded as animals might be still adjusting to the experimental settings. Figure 3.3.11 shows the latency, recorded as time to fall from the accelerated rod, for each control and gene therapy group. In this experiment only WT group showed statistical difference in latency compared to *Gaa -/-* animals (*$p < 0.05$, +/- SEM, n=8 for WT and n=6 for *Gaa -/-* mice). PGK-SSP-NewTagI-II and PGK-GAA native did not outperformed *Gaa -/-* animals (n=5 for PGK SSP NewTagI-II, n=4 PGK GAA native, n=6 for *Gaa -/-* animals, +/- SEM). In addition, no significant difference in latency of functional test was seen between PGK SSP NewTagI-II and PGK GAA native treated animals.
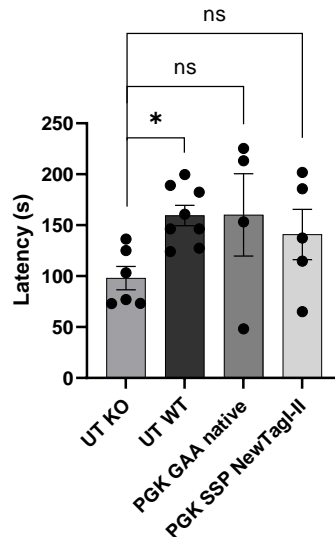
*Figure 3.3.11. Muscle coordination test.* *A rotarod test was undertaken to determine functional correction of Gaa -/- phenotype mice. Each animal was a subject to five runs separated with 3 minutes of cooldown to rest. Data from two initial runs was discarded and only remaining 3 consecutive runs counted toward functional analysis. UT KO = untreated Gaa -/- mice, UT WT = untreated wild type animals. KO UT group (n=6) showed statistically significant lower latency compared only to WT group (n=8), and no significant difference to PGK-SSP-NewTagI-II (n=5) and PGK-GAA native (n=4) (\*p<0.05, +/-SEM). Data analysed using Two-way ANOVA for repeat measures, corrected for multi comparisons using Bonferroni's statistical hypothesis test.*

In addition to functional correction in skeletal muscle test an echocardiography was performed, by our collaborators at Centre for Advanced Biomedical Imaging at UCL, and raw data collected for each animal to determine cardiac muscle improvement i.e., dimensions and functions including left ventricle mass (LV) (Figure 3.3.12 A), cardiac output (CO) (Figure 3.3.12 B), ejection fraction in % (EF) (Figure 3.3.12 C), left ventricular posterior wall diastole (LVPW) (Figure 3.3.12 D), left ventricular posterior wall systole (LVPW)(Figure 3.3.12 E), intraventricular septum diastole (IVS diastole) (Figure 3.3.12 F) and interventricular septum systole (IVS systole) (Figure 3.3.12 G). Out of all measurements taken only one showed significant and one close to significant difference between PGK-SSP-NewTagI-II GAA treated group and UT *Gaa -/-* including ejection fraction Figure 5.13 B \*p<0.05 and left ventricular posterior wall systole seen on Figure

5.13 E p=0.09. Remaining measurements showed no significant differences between *Gaa -/-* animals, WT, and gene therapy treated groups. In addition, there was also no significant difference shown between PGK SSP NewTagI-II and PGK GAA native groups.

Interestingly Piras *et al.* (Piras et al., 2020) in 6 months *in vivo* Pompe study showed clear differentiation in cardiac muscle dimension and functions using the same *Gaa -/-* mice model and Wild type group. We hypothesised that perhaps undertaking end points at 3 months, with animals reaching only 21 weeks of age across all experimental groups, was too soon to allow to generate a sufficient deterioration to cardiac muscles in *Gaa -/-* mice and clear separation from WT and gene therapy groups. Perhaps also a small number of animals in each experimental arm did not provide significant separation of measured functions of cardiac muscle.
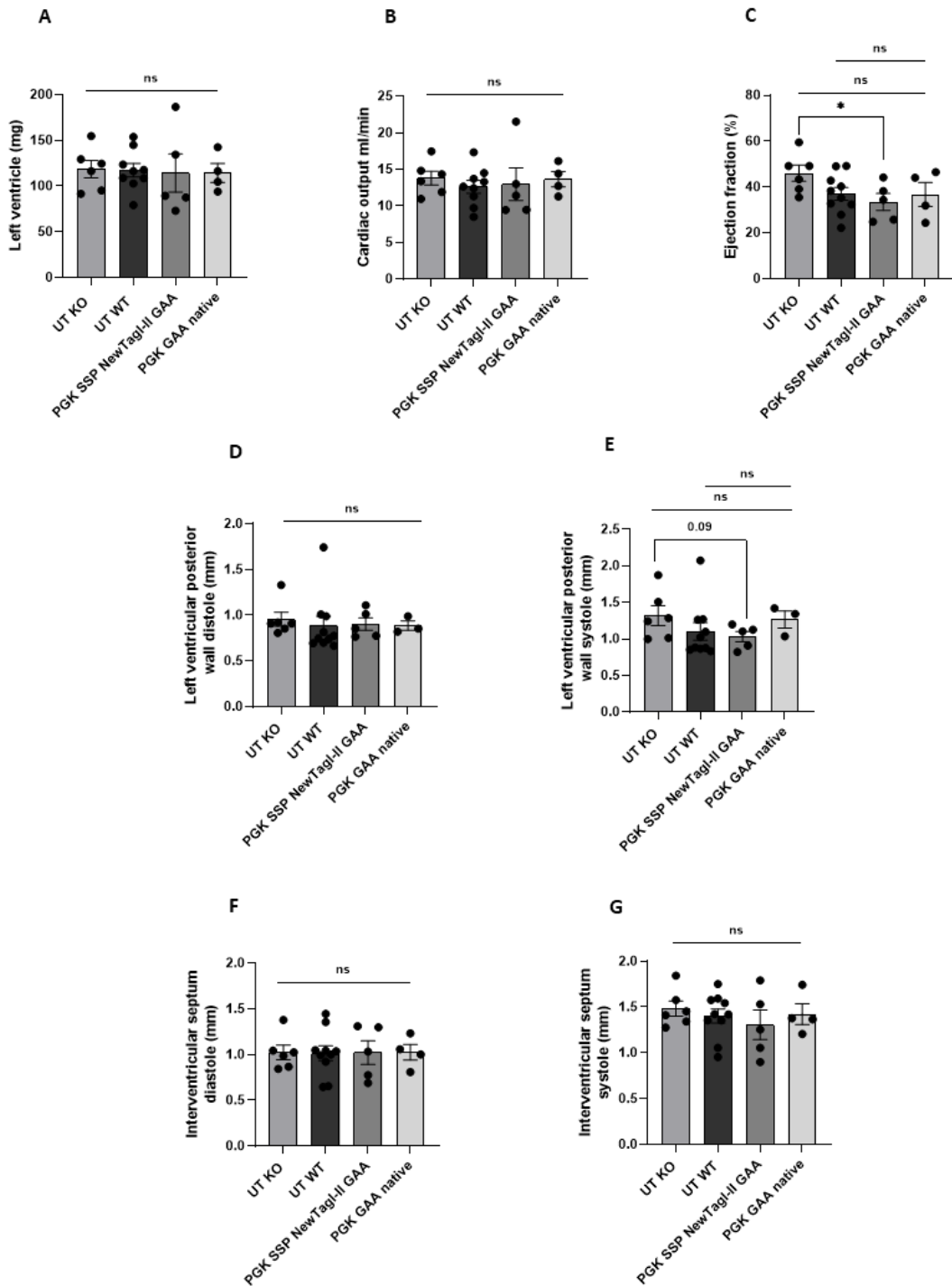
***Figure 3.3.12 Echocardiography in experimental animals.*** *Individual animals before final termination underwent an echocardiography analysis. The analysis focused on cardiac muscle's dimension and functionality of the organ. Between the experimental groups the data showed no difference in **A)** left ventricle mass (LV) and in **B)** cardiac output (CO). In **C)** ejection fraction % (EF) showed a statistical difference between UT KO and PGK SSP NewTagI-II animals (p<0.05). Furthermore, **D)** left ventricular posterior wall diastole (LVPW) showed again no significant difference in-between groups, however **E)***

*left ventricular posterior wall systole (LVPW) measurements showed p=0.09 significance between UT KO and PGK SSP NewTagI-II animals. In addition, neither of **F)** the intraventricular septum diastole (IVS diastole) nor **G)** interventricular septum systole (IVS systole) measurements indicate major difference between experimental groups. No significant difference was observed in any of the measurements performed between PGK SSP NewTagI-II and PGK GAA native groups. (All measurements were performed on UT KO n=6, UT WT n=8, PGK SSP NewTagI-II n=5 and PGK GAA native n=4 animals, +/-SEM, Kruskal-Wallis H test).*

### 3.3.3 Discussion chapter

In this chapter *in vivo* gene therapy data from transduced HSCs into Gaa -/- mice was interrogated.

Sweeney *et al.* studying LV integration into host genetic material concluded that the total size of transgene and transgene promoter greatly impacted the ability of lentivirus to unpack and process with reverse transcription of its genetic material leading to lower VCN in consequence (Sweeney & Vink, 2021; Terwilliger et al., 1989). Therefore, we decided to substitute the LCR-EFS1 tandem promoter for the PGK promoter alone and in doing so reduced the promoter and transgene cassette size by 2 kb. This substitution led to a substantial increase in VCN in transduced and transplanted HSCs. Analysis of activity in blood plasma of all arms of the *in vivo* experiment showed variety of levels of GAA activity with particularly high levels of enzymatic activity seen in SSP-NewTagI-IIGAA cohort (Figure 3.3.7). However, this did not translate into greater glycogen reduction in tested tissues compared to tissues of animals treated with native form of GAA. The significance in GAA uptake (Figure 3.2.7) and reduction of glycogen (Figure 3.2.8) seen in SSPNewTagI-IIGAA treated *Gaa -/-* myofibers seemed to be inadequate *in vivo*. This might suggest that NewTagI-II motif is not optimally design to facilitate sufficient plasma membrane translocation and lysosomal targeting for GAA protein to deliver a full cross correction. In addition, the lack of full correction can also be associated to the NewTag's plasma counterpart/receptor. This potential receptor is yet unknown, and its distribution/expression level on the plasma membrane is also

unknown. Furthermore, binding affinity of NewTag to its potential plasma receptor remains concealed. It is plausible that the partial correction observed in vivo can be the result of many intertwined factors.

However, it is very reassuring that our LV delivery platform bearing *GAA* transgene (+/- NewTag) showed a significant decrease in glycogen across all tissues compared to *Gaa-/-* untreated animals. This finding follows previous studies performed by Stock *et al.* and Piras *et al.* showing high correction of PD phenotype in animal model ( Piras et al., 2020; Stok et al., 2020). Additionally, brain tissue that is inaccessible for ERT and AAV liver targeted gene therapy delivery, became reachable for delivering PD correction when using transduced and transplanted HSCs. Glycogen reduction in brain tissue shown on Figure 3.3.8 B reflect changes seen in previous studies on PD.

The levels of GAA activity across tested tissues are debatable and can be associated to incorrect flushing, leaving rhGAA presented in blood, amalgamated with a tissue adding to an increased overall level of GAA activity measured from homogenised organs. In particular liver and lungs showed high GAA signal in homogenised fractions. However, giving that these two organs are highly vascularised tissues  (Chamarthy et al., 2018; Lorente et al., 2020) it is plausible that the elevated GAA activity observed can also be associated to the high number of hematopoietic cells present, adding to the insufficient flashing. However, this would need to be further evaluated, if possible.

Furthermore, cardiac muscle echocardiography unveiled some improvements in cardiac functions and output in both gene therapy cohorts compared to untreated *Gaa -/-* mice, however there was no significant difference between each gene therapy construct bearing groups and control (Figure 3.3.12). Also, in comparison to previous study conducted by Piras *et al.* (Piras et al., 2020) the correction in our study seems less pronounced. However, Piras *et al.* assessed correction of PD at 6 months old *Gaa-/-* animals compared to 3 moths used in this study.  Similarly, rotarod values across GT treated groups did not show significant difference to those of *Gaa-/-* untreated animals. Yet again, in study by Piras *et al.* GT treated  *Gaa-/-* animals showed difference in muscle strength and coordination compared to *Gaa-/-* untreated animals, but yet again those animals were examined after 6 months of GT in contrast to 3 months in our study. Also, rotarod might have not been the most favourable method to determine muscle

correction/strength and alternative devices i.e., grip string test and/or treadmill would be more informative, however were unavailable to undertake this in the timeframe of the experimental endpoint.

In both cases, cardiac muscle echocardiography and rotarod presented by Piras *et al.* an additional time of 3 months perhaps allowed GT treated *Gaa-/-* to produce clearer correction compared to untreated *Gaa-/-* mice. However, this would need to be further investigated in additional *in vivo* studies; however, the time frames of this PhD study would not permit it.

## 3.4  Investigation of the muscle homing peptide (NewTag motif) uptake mechanism and plasma membrane counterpart.

Despite several studies (Gao et al., 2014; Tsoumpra et al., 2019) using MHP as a therapeutic agent for cellular uptake, its cellular point of entry remains yet unknown. There are several mechanisms of cell entry with some being passive i.e. entry of small molecules/chemicals and other being more selective and specialised based on the interaction between plasma membrane receptor and its ligand (Yang & Hinner, 2015).

Each individual mammalian (eukaryotic) cell expresses a vast number of plasma membrane receptors in order to detect and respond to an array of physical and chemical ligands (stimuli) from extracellular environment (Vauquelin & Packeu, 2009). As intracellular/extracellular interactions are extremely complex and specific, for that, the number and specification/specialisation of receptors reflect the need to responds specifically and quickly to each stimulus (Chen et al., 2017). There are about 25 different families of receptors with each family encompassing a vast number (up to a thousand) of plasma membrane proteins (Amarante-Mendes et al., 2018). Several families of receptors upon their activation propagate a signal (signal transduction) throughout plasma membrane into cytoplasm, activating a specific signalling pathway to exhibit a precise respond to a particular activating ligand (Rosenbaum et al., 2009). Alternatively to signal transduction,  numerous of receptors allow to bind, capture and enclose macromolecules (ligands) within membrane-bound vesicles that invaginate and pinch off the plasma membrane in a process called endocytosis (Aguilar & Wendland, 2005).

Cells use several different mechanisms for endocytosis including micropinocytosis, phagocytosis, caveolae-dependent endocytosis, clathrin-facilitated uptake and neither caveolae nor clathrin-mediumted endocytosis (Kumari et al., 2010). Endocytosis triggered by activation of a specific receptor is called receptor mediumted endocytosis (RME) and it is a critical for i.e. obtaining essential nutrients, immunological responds to pathogens, (Alberts et al., 2015), regulating developmental signalling (Kaksonen & Roux, 2018) and engaging  in turnover of plasma membrane constituents (Berro, 2019). For

instance, earlier described CI-M6PR is a clathrin-mediumted endocytic receptor (Lin et al., 2004). Upon CI-M6PR activation by its ligand, assembly of endocytic machinery begins with first recruitment of core adaptor proteins AP2 to plasma membrane (Alberts et al., 2015). AP2 becomes phosphorylated and acts as the major hub for the interaction between membrane and clathrin sheets (Hansen & Nichols, 2009). Following the clathrin coat assembly and the plasma membrane invagination, GTPase dynamin proteins construct a "collar" over a neck of the pit and a clathrin coated vesicle detaches from plasma membrane. Further, the fate of the vesicle and its cargo is determined by the original signal produced upon its receptor activation and, after endosomal sorting, it might target trans-Golgi network (TGN) or Lysosomes (Alberts et al., 2015).

From a therapeutic point of view, the RME is a very promising delivery pathway for treating LSDs as the majority of vesicles/early endosomes are targeting lysosomes for degradation of its cargo (Alberts et al., 2015). ERT based on GAA recombinant protein, through CI-M6PR and clathrin-mediumted endocytosis (CME) (Mettlen et al., 2018), is delivered to lysosomes exhibiting its therapeutic effects by reducing glycogen burden in treated patients (Koeberl et al., 2011b). Based on aforementioned studies by Gao *et al.* and Tsoumpra *et al.* (Gao et al., 2014; Tsoumpra et al., 2019) MHP shown ability to enter through plasma membrane and delivers its cargo into muscles cells. Therefore, MHP entry through endocytosis-related process is highly probable but the exact mechanism it is yet to be discovered.


In order to interpret and explain better our *in vivo* results, I performed a series of studies to elucidate the potential mechanism of entry of NewTag motif and to study its interaction with plasma membrane proteins to identify a putative plasma membrane receptor.

**3.4.1 Dissecting the pathway of entry of NewTag motif in cells.**

There are several different families of plasma membrane receptors through which signals/molecules enter across plasma membrane into cells and exercise their primary objectives (Amarante-Mendes et al., 2018; Yang & Hinner, 2015). Out of all these families of receptors and internalisation mechanisms, in means of gene therapy delivery to Pompe patients, endocytic receptors are of particular interest. Jirka *et al.* (Jirka et al., 2018) studying cyclic peptide delivery in mouse model of Duchenne Muscular Dystrophy used several different conditions to discriminate between different families of endocytosis receptors and/or mechanisms of entry. Following this lead, I set up a panel of inhibitors (Table 3.4.1) to determine if *Gaa -/-* C2C12 myofibers, treated with each individual condition separately and 1μM NewTagI-II GAA concentrated medium in presence of M6PR inhibitor, show difference in GAA activity. In addition, *Gaa -/-* C2C12 in presence of M6PR inhibitor were cultured with 1μM NewTagI-II GAA and also plain K562 concentrated medium to establish basal level of GAA activity.

**Table 3.4.1. Pharmacological and physical inhibitors.**

| Condition | Mechanism |
|---|---|
| 4°C | All energy-dependent uptake |
| Sodium azide | ATPase inhibition |
| Chlorpromazine | Inhibition of clathrin-mediated uptake |
| Chloroquine | Disruption of endocytic vesicles |
| Genistein | Obstruction of caveolin-mediated uptake |
| Fucoidan | Inhibition of scavenger receptor-mediated uptake |
| Dextran sulfate | Interruption of scavenger receptor-mediated uptake |

After 24hrs of treatment, *Gaa -/-* C2C12 were collected, and GAA activity tested and presented on Figure 3.4.1. Samples collected from 4°C and sodium azide showed significant difference in GAA activity to cells cultured in presence of NewTagI-II-GAA and M6P (** $p<0.01$ and *$p<0.05$ respectively, +/-SEM, n=5) indicating that MHP exhibits energy dependent mechanisms of entry. Similarly, chlorpromazine and chloroquine, both inhibiting clathrin-dependent endocytosis, impacted significantly GAA activity compared to control sample narrowing down the information on the potential

mechanism of entry (**p<0.01 and p=0.07 respectively, n=5, +/-SEM). In contrast, *Gaa -/-* C2C12 cells exposed to either genistein, fucoidan or dextran sulphate shown no significantly lower level of GAA activity compared to control sample suggesting that neither caveolin nor scavenger-dependent endocytosis was employed for internalisation of NewTagI-II GAA recombinant protein.
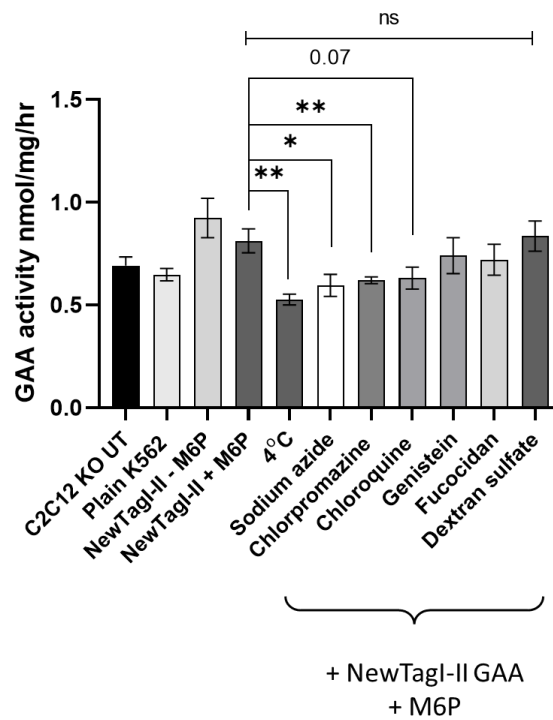


***Figure 3.4.1. Investigation of NewTagI-II mechanism of entry in cells.*** *Gaa -/- C2C12 myofibers were exposed to 1µM of NewTagI-IIGAA (or plain concentrated K562 cells medium) and presence of 5mM M6PR inhibitor (unless otherwise stated), and for 24 hr in presence of pharmacological endocytosis inhibitors or low temperature. Control samples for this assay were cells exposed to 1µM of NewTagI-IIGAA + M6P only. After 24hr incubation, all samples were harvested, cell pellets lysed to measure GAA activity. The exposure to4ºC and Sodium azide conditions showed significantly lower enzymatic activity compared to control samples (**p<0.01 and *p<0.05 respectively, +/-SEM, n=5). Furthermore, Chlorpromazine significantly inhibited GAA activity in treated samples compared to the control (*p<0.05, n=5, +/-SEM), followed by Chloroquine with a p-value*

*of 0.07 (+/-SEM, n=5). Genistein, Fucoidan and Dextran sulfate did not significantly influence GAA level in treated Gaa -/- C2C12 myofibers (+/-SEM, n=5, one-way ANOVA with Tukey's post hoc test was used to analyse experimental groups).*

**3.4.2 Investigate a cathrin-mediumted endocytosis pathway.**

Following the above findings, a clathrin-depended endocytosis pathway was further investigated to understand the mechanism of entry of NewTag motif into cells.

Clathrin-mediumted endocytosis (CME) is a rapid process involving the activation and assembly of core proteins to a plasma membrane to internalise and transport a receptor and its cargo into endocytic vesicles; this process might occur extremely rapidly and may only last a matter of minutes from activation (Kumari et al., 2010). After extensive study of available literature, I proposed to examine the level of phosphorylated form of AP2 (AP2-P) that is part of the core endocytic proteins and can be used as a marker of activated CME.

**3.4.2.1 Determination of AP2 phosphoprotein formation in *Gaa -/-* C2C12 cells in response to NewTagI-IIGAA from concentrated medium.**

Following on several publications exploring CME pathway (Rappoport & Simon, 2003; Rennick et al., 2021), I selected two timepoints: 5- and 10-mins as endpoints to check AP2-P formation in *Gaa -/-* C2C12 cells upon exposure to NewTagI-IIGAA concentrated medium.

Briefly, *Gaa -/-* C2C12 myofibers were exposed to 1µM NewTagI-IIGAA concentrated or plain K562 medium of equal volume. In both conditions 5mM M6PR inhibitor was added to the culture. *Gaa -/-* C2C12 untreated myofibers were also used a baseline control for AP2-P level. After 5- and 10-minutes, cells were washed, harvested, and immediately placed on dry ice to stop intracellular process and preserve, if any, activated AP2-P

protein. Samples were then lysed using RIPA buffer supplemented with phosphatases and proteases inhibitors.

Western blot analysis using anti- AP2-P immunoglobulin revealed the presence of AP2-P in samples exposed for 5 and 10 minutes both to NewTagI-IIGAA (Figure 3.4.2 A) and plain K562 (Figure 3.4.2 B) compared to untreated *Gaa -/-* C2C12 cells. *Gaa -/-* C2C12 cells exposed to NewTagI-IIGAA for 10min showed a stronger and sustained activation signal compared to the one exposed to plain K562 medium. While re-probed membrane with AP2 antibodies revealed no differences in total AP2 level between treated samples and control untreated *Gaa -/-* C2C12 cells (Figure 3.4.2 A and B middle line).
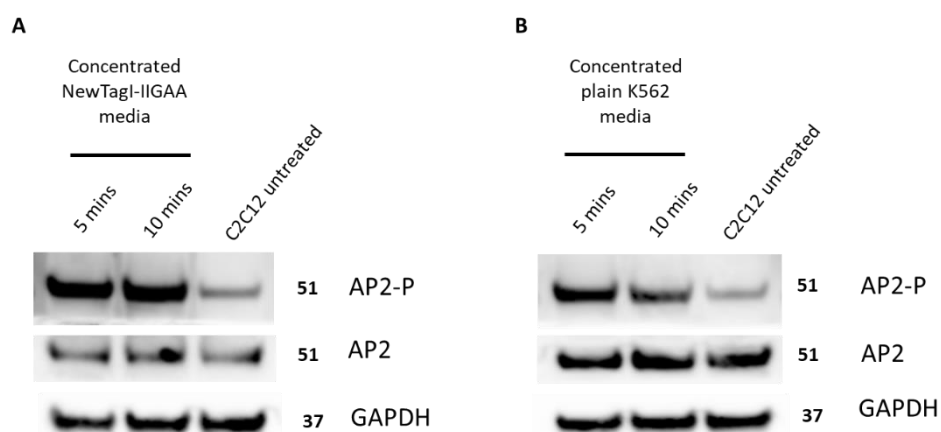


***Figure 3.4.2. AP2-P formation and endocytosis activation in Gaa -/- cells upon exposure to NewTagI-IIGAA.*** *Western blot **of** Gaa -/- C2C12 myofiber lysates exposed for 5 or 10 minutes to **A)** concentrated medium containing 1µM NewTagI-IIGAA or **B)** plain K562 medium. Untreated Gaa -/- C2C12 myofiber lysates were used as control for basal level of AP2/AP2-P signal. Top blot: AP2-P signal; mid blot: AP2 signal; bottom blot: GAPDH loading control (n=1).*

**3.4.2.2 NewTagI-II tag protein fused to Myc and His motifs only.**

In order to expose *Gaa -/-* C2C12 cells purely to MHP sequence and avoid interference of medium components and cellular metabolites in the study of CME, I cloned fragment containing SSP-NewTagI-II in tandem to Myc-His tag into a pCCL LV plasmid (Figure 3.4.3

A) alongside a fragment expressing SSP-Myc-His only, which served as control (Figure 3.4.3 B). This tool was created to detect the AP2-P signal triggered by NewTagI-II binding only to *Gaa -/-* C2C12 cells. His tag (amino acids structure made of 6 Histidine residues HHHHHH) was employed to purify the fragments through chromatography HiTrap column from cell culture medium, and Myc structure (amino acids composition EQKLISEEDL) was used to visualise the fragments on immunoblotting membrane.

A



B
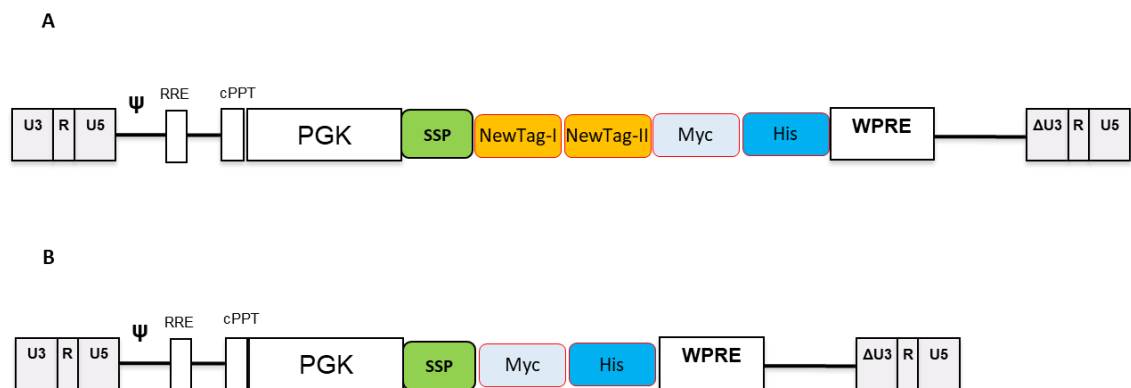


***Figure 3.4.3. Design of lentiviral vector inserts containing Myc and His tag protein.*** *Structure of second generation pCCL lentiviral vector backbone was employed to clone in **A)** SSP-NewTagI-II-Myc-His fragment or **B)** SSP-Myc-His short piece which served as a control peptide.*

### 3.4.2.3 AP2-P level in *Gaa -/-* C2C12 cells exposed to purified NewTagII-Myc-His protein

Chromatography HiTrap column was used to purify both constructs NewTagI-II-Myc-His and Myc-His from cell culture medium. Elution was run on WB to determine success rate of prefinalisation both tag proteins. Anti-Myc immunoblotting revealed that NewTagI-II-Myc-His protein is 12 kDa, and ran higher than Myc-His 7 kDa, as expected (Figure 3.4.4).
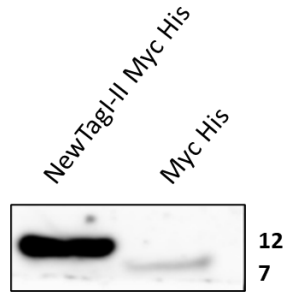
**Figure 3.4.4. Detection of Myc tag containing proteins purified from medium by HisTrap column.** *A total of 30µl of each eluted fraction containing NewTagI-II-Myc-His or Myc-His was ran on Western blotting. Primary rabbit anti-Myc 1:1000 follow by secondary any rabbit HRP-conjugated 1:1000.*

NewTagI-II-Myc-His on WB shown higher concentration in given volume than Myc-His eluted from HiTraps columns, therefore, to avoid any bias caused by higher accumulation of NewTagI-II-Myc-His this construct was diluted 10 times to reach similar concentration to Myc-His tag protein.

Differentiated *Gaa -/-* C2C12 myofibers were treated with 10, 30, 50 and 80µl of each of the eluted tag proteins. After 5 minutes cells were washed, harvested and cell pellets placed on dry ice. Further, samples were lysed using RIPA buffer supplemented with phosphatases and proteases inhibitors.

Immunoblotting with anti AP2-P revealed substantial increase of AP2-P signal in treated *Gaa -/-* C2C12 cells with NewTagI-II-Myc-His in a dose dependent manner over untreated control sample of *Gaa -/-* C2C12 cells (Figure 3.4.5 A). In contrast, cells treated with Myc-His tag protein showed no increase in AP2-P over untreated samples (Figure 3.4.5 B). Stripped membranes and reblotted with anti AP2 protein showed no differences between treated conditions and untreated samples for NewTagI-II-Myc-His and Myc-His (Figure 3.4.5 A and B respectively).

**A**

NewTagI-II Myc His

μl  10   30   50   80   UT

51 AP2-P

51 AP2

37 GAPDH

**B**

Myc His

μl  10   30   50   80   UT

51 AP2-P

51 AP2

37 GAPDH
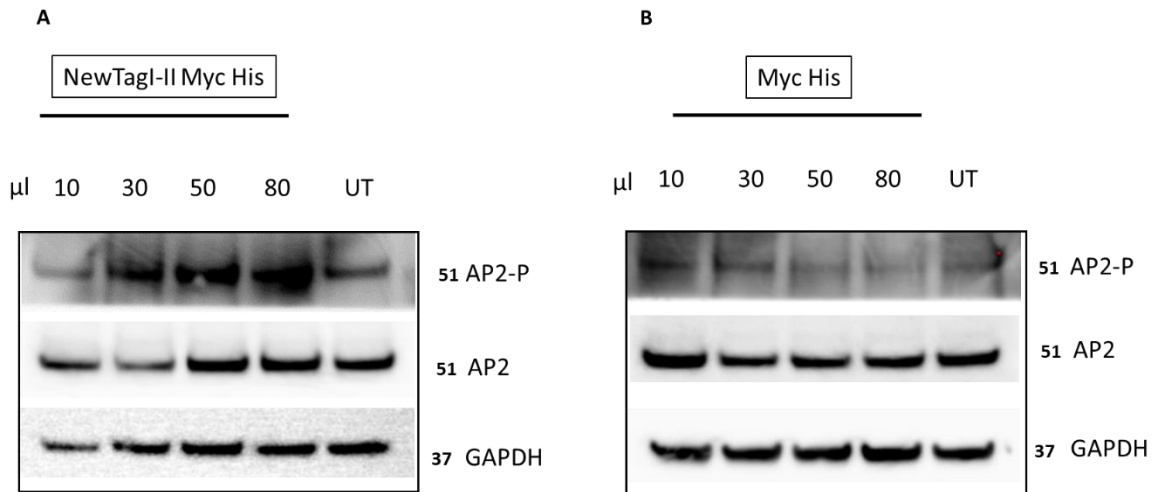
***Figure 3.4.5. AP2-P signal detection in samples treated with NewTagI-II-Myc-His vs Myc-His tag protein. A)*** *Gaa -/- C2C12 cells exposed for 5 minutes to series of volumes (10,30,50 and 80μl) of NewTagI-II-Myc-His. 20μg of total lysed proteins were loaded per each sample/volume used in Western blotting. Anti-AP2-P immunoblotting showed gradual increase in phosphoprotein signal (top blot) compared with control sample.* ***B)*** *Control Myc-His tag protein was also used for 5 minutes in series volumes (10, 30, 50 and 80μl) to treat Gaa -/- C2C12 cells. 20μg of total lysed proteins per each sample were blotted on Western blot. Primary rabbit anti-Myc 1:1000 follow by secondary anti rabbit HRP-conjugated 1:1000 were used to visualise Myc motif containing protein. (n=1).*

Findings from Figure 3.4.5 A demonstrated that NewTagI-II domain has an interaction with a plasma membrane protein and activate an endocytic response upon binding. To investigate the reason why NewTagI-II GAA construct did not exert the desired increase in uptake in the *in vivo* study, I decided to study whether NewTagI-II binding is potentially reduced upon combination with GAA protein.

Using ConA chromatography column (as described in chapter 2.9.1 materials and methods), NewTagI-IIGAA recombinant proteins were purified from medium of transduced K562 stable cell line and used at the concentration of 0.25, 0.5, 1, and 2 μM to treat *Gaa -/-* C2C12 myofibers in the presence or absence of M6PR inhibitor. Along this set of samples, differentiated *Gaa -/-* C2C12 cells were also exposed to 10, 30, 50

and 80 μl of NewTagI-II-Myc-His in the presence or absence of M6P. All cell samples were treated as previously for 5 minutes and then washed, harvested and cell pellets lysed to determine AP2/AP2-P levels by western blots.

Western blot data showed an increase in AP2-P signal in cells treated with NewTagI-II-Myc-His tag protein in presence of 5mM M6P compared to untreated cells (Figure 3.4.6 A). In contrast cells treated with NewTagI-II GAA construct in presence of 5mM M6P showed no distinguishable difference in AP2-P signal compared to untreated cells (Figure 3.4.6 B). However, cells treated with NewTagI-II GAA construct but without M6PR inhibitor showed substantial increase in AP2-P signal compared to untreated samples (Figure 3.4.6 B'). In addition, cells treated with NewTagI-II-Myc-His tag without M6P inhibitor showed also increase in AP2-P signal (Figure 3.4.6 A'). Stripped and reblotted membranes with anti-AP2 immunoglobulin showed no difference in signal intensity between cells treated with NewTagI-II-Myc-His +/-M6P, NewTagI-II GAA +/-M6P and untreated samples (Figure 3.4.6 A, A', B and B').

WB signal from AP2-P across all 4 blots was then quantified and corrected by GAPDH signal. Figure 3.4.6 C and C', representing the quantification of Figure 3.4.6 A and A' immunoblotting, respectively, showed that NewTagI-II-Myc-His regardless of a presence or absence of M6P inhibitor allowed for substantial phosphorylation of AP2 protein. In contrast, Figure 3.4.6 D and D', showing quantification of Figure 3.4.6 B and B' respectively, indicated that cells pre-treated with M6P inhibitor follow by exposure to NewTagI-I- GAA construct showed little to none AP2-P signal (Fig 3.4.6 D and B), and on the contrary, cells exposed to NewTagI-II-GAA protein and cultured without M6P inhibitor generated AP2-P signal compared to untreated cells (Figure 3.4.6 B' and D').

This data suggested that NewTagI-II might have reduced its effect on construct internalisation when combined with GAA protein. This finding further elucidated a lack of additive uptake stimulation seen in SSPNewTagI-II-GAA treated group in *in vivo* experiments.
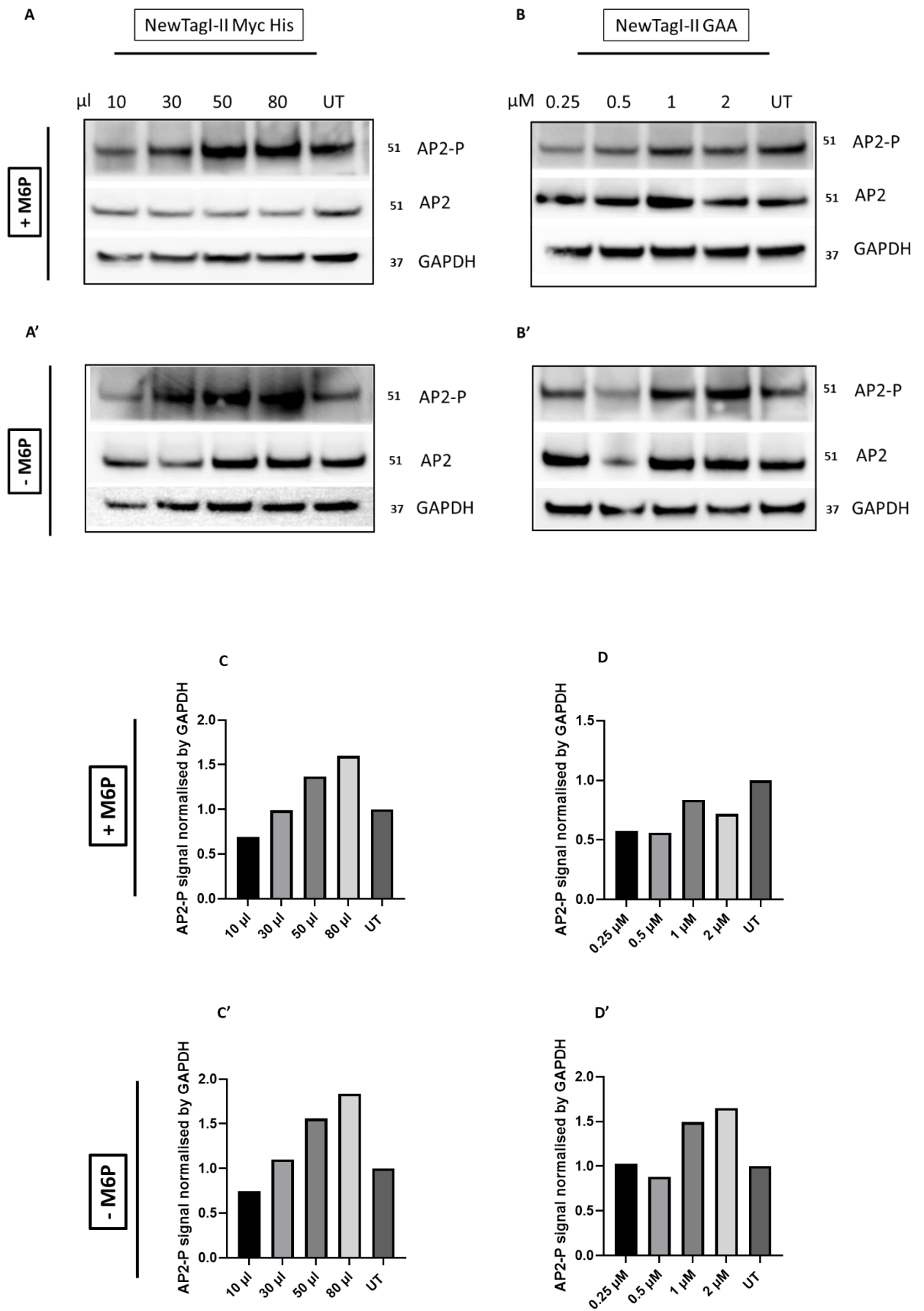
**Figure 3.4.6. Comparison of NewTagI-II-Myc-His and NewTagI-II-GAA ability to trigger the phosphorylation of AP2.** *Gaa -/- C2C12 cells were subject to incubation with*

*NewTagI-II-Myc-His peptide **A)** in the presence or **A')** absence of M6PR inhibitor, or with NewTagI-II-GAA purified protein **B)** with or **B')** without M6PR inhibitor Cells were exposed to 5mM M6P for 1 hr prior to addition of the tag proteins in 37°C and 5% $CO_2$ incubator. Original stock of NewTagI-II-Myc-His was diluted 10x, and 10, 30, 50 and 80µl of the diluted stock was added to cells followed by 5 min incubation. While NewTagI-II-GAA peptide was used at 0.25, 0.5, 1 or 2 µM concentration. **C, C')** Quantification of AP2-P band brightness normalised by GAPDH from Figure A and A', respectively. **D, D')** Quantification of AP2-P bands brightness normalised by GAPDH from figure B and B', respectively. Quantification of Western blot bands' intensity was carried by ImageJ software. (n=1)*

Despite that NewTagI-II motif combined with GAA protein did not provide a significant and instant activation of AP2-P protein, its action might still be biologically relevant over longer period of time as seen in 24hr exposure *in vitro* uptake experiments.

### 3.4.3 RNA-seq analysis.

CME is a very quick process through which cells respond immediately for incoming stimuli from extracellular space, from cell-to-cell interaction and/or foreign invaders i.e., viruses. Receptors on cell surface responding to incoming stimuli and activating CME are also very quickly activated, enclosed into clathrin-coated vesicle and recycle back to the plasma membrane upon merging a transport vesicle with early endosomes (Doyon et al., 2011; Geljic et al., 2021). The turnover of plasma receptors and CME core protein and adaptors, exposed to a constant stimulation, is great and in a short period requires new synthesis of specific components of the CME system (McPherson et al., 2009; Metzler et al., 2001; Ritter et al., 2011). This follows particular gene activation and transcription of specific depleted proteins. These transcripts can serve as a snapshot of biological activities and shed some light on undergoing processes inside cells at that given time (Huang et al., 2021; Janiszewski et al., 2021; Smallets & Kendall, 2021; Yoko-o et al., 2021). However, the timing to take the snapshot is limited. Therefore to determine gene expression of CME related receptors, core proteins, adaptors, and any

recycling pathways, we decided that in the first instance, we would investigate an early exposure time of 6hrs of C2C12 *Gaa-/-* to NewTagI-II tag protein and explore upregulated gene set only as a primary response to depleted level of CME proteins (Anjum et al., 2016; Balliu et al., 2019; Boyce et al., 2020; Bryois et al., 2017).

### 3.4.3.1 Gene set enrichment analysis.

Three biological replicates were set up using differentiated *Gaa -/-* C2C12 cells in 6 well plate with two conditions per each replicate consisting of untreated cells and cells treated with 300 µl of the NewTagI-II-Myc-His tag for 6 hrs.

After incubation, cells were harvested, washed and processed to total RNA extraction using Qiagen RNeasy extraction kit. Extracted RNA was immediately processed for total RNA sequencing by Genomics department at University College London (UCL) (https://www.ucl.ac.uk/child-health/research/genetics-and-genomic-medicine/ucl-genomics) using Ilumina sequencing kit at NextSeq2000 instrument with 100 cycles and output of 66.7 million reads in total. Analysis of raw data, gene alignments including differential expression analysis and separation between upregulated and downregulated genes were also processed by UCL Genomics.

A quality control of acquired data was perform by checking variability within the experimental groups. The main variability within the experiment was expected to come from biological differences between the samples. This variability was checked by using Variance Stabilizing Transformation (VST) (Hafemeister & Satija, 2019) which performed a hierarchical clustering of the entire sample set and led to the generation of cluster dendrogram (Figure 3.4.7). An Euclidean distance was computed between samples, and the dendrogram was constructed upon the Ward criterion (Ward's minimum variance criterion minimizes the total variance within-cluster (Strauss & von Maltitz, 2017). Constructed dendrogram showed clear clustering to group replicates and separation between biological conditions: treated and untreated samples.
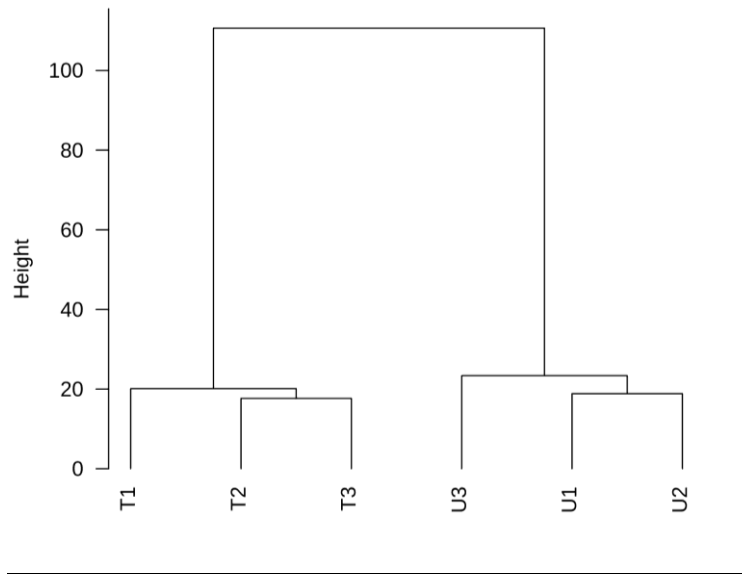
***Figure 3.4.7. Cluster dendrogram***. *Clear clustering withing biological replicates and separation between two tested conditions treated (T1-3) and untreated (U1-3) using Ward criterion clustering method indicated good quality of produced samples and provided with statistical power to the RNA-seq results in differential gene expression analysis.*

The VST was performed by UCL Genomics as a part of quality control process. RNA-seq cross-examined a total of 17899 genes, with 7426 genes showing upregulation and 10373 genes found downregulated and/or neutral (no change) in treated vs untreated clusters. Significant difference between treated cluster vs untreated on at least $p<0.05$ threshold was shown by 3562 upregulated genes and 3551 downregulated genes. $Log_2$ fold change vs. $-log_{10}$ adjusted P value was plotted on volcano plot to depicture a distribution of up and down regulated (with at least $p<0.05$ value) genes (Figure 3.4.8 A)

Further analysis of gene enrichment in treated samples consisted of separation of endocytosis related genes from a total pool of 3562 upregulated genes (Figure 3.4.8 B). KEGG mouse library comprised of 185 genes related to endocytosis signalling pathway was used (https://www.genome.jp/kegg-bin/show_pathway?mmu04144). Using a Molecular Signature Database platform (MSigDB, http://www.gsea-msigdb.org/gsea/msigdb/index.jsp) the set of enriched (with minimum of $p<0.05$) genes

were checked against KEGG endocytosis signalling library, revealing a 62 genes upregulated and linked to the endocytic pathway (Figure 3.4.8 C).
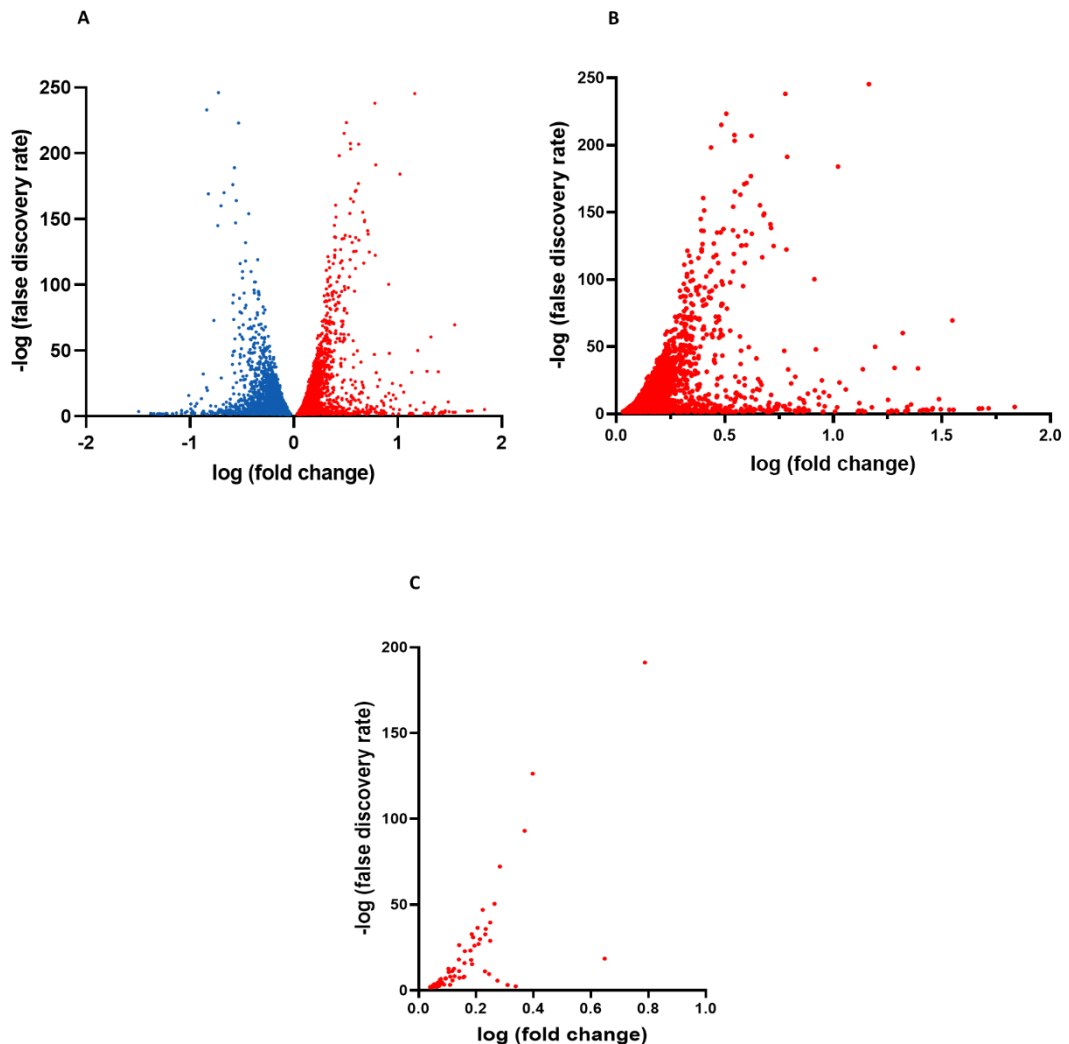


***Figure 3.4.8. RNA-seq gene set enrichment analysis****. Volcano plot represented statistical significance (P value) defined also as -log (false discovery rate) vs degree of change also called fold change. This scatterplot provided a visual detection of genes (upregulated in* red *and downregulated in* blue) *with large fold changes treated vs untreated samples that are also statistically significant and might be the most biologically significant genes in given experiment. **B)** Focus on upregulated genes presented with at least p<0.05 treated vs untreated conditions. **C)** Volcano plot with only endocytosis relevant genes presented in enriched upregulated genes with at least p<0.05.*

Out of the 62 genes involved in the endocytosis pathway a total of 4 plasma membrane receptors were shown in the MSigDB search to be significantly upregulated in treated cluster compared to untreated samples. All 4 receptors are capable of activation of CME: Chemokine receptor 4 (CXCR4) (DeNies et al., 2019), Low density lipoprotein receptor (LDLR) (Goldstein & Brown, 2009), Platelet derived grow factor receptor α (PDGFRa) (Rakhit et al., 2000) and Scavenger receptor B1 family (SR-BI encoded by SCARB1 gene) (Lenahan et al., 2019) (Figure 6.9 A and B). This last one drew my attention because it supposed to not be a CME receptor, however it was shown to be an endocytic receptor (Zani et al., 2015). After intense search through literature, it transpired that SR-BI has a spliced isoform known as SR-BII, which differs from SR-BI only in the cytoplasmic carboxyl-terminal tail, with no difference in the extracellular compartment. Also most importantly, SR-BII is located to clathrin-coated pits on plasma membranes and activates CME (Eckhardt et al., 2006).

Furthermore, the gene enrichment analysis indicated upregulation in genes coding for light chain and heavy chain of clathrin complex CLTA/CLTC and AP-2 adaptor protein and many genes related to endosomal/lysosomal transporting and endosomal recycling, shown in Figure 3.4.9 C and also listed in Table 3.4.2, supporting CME. These findings further suggested that NewTagI-II tag abilities to target endosomal/lysosomal vacuolar system.
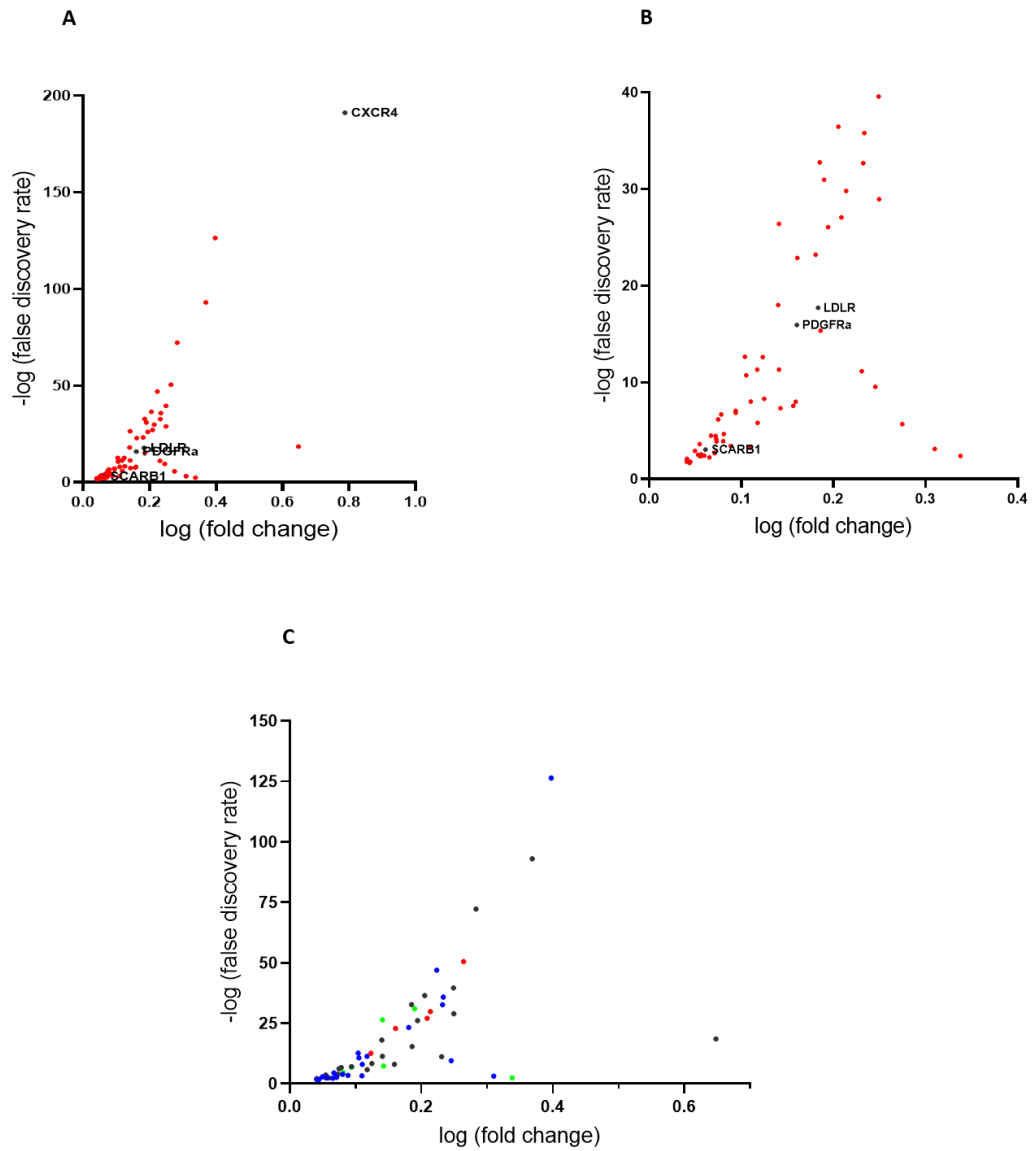
**Figure 3.4.9. Endocytic receptors in enriched set of genes. A, B)** *4 plasma membrane receptors were identified in the pool of endocytose related genes including CXCR4, LDLR, PDGFRa and SCARB1(SR-BI).* **C)** *Division of endocytosis related genes into four functional/structural groups:* blue adaptors and core proteins, *black membrane trafficking/ endosome/ lysosome/ recycling,* green plasma membrane scaffolding/ cytoskeleton, *and* red E3 ubiquitin- protein ligase.

**Table 3.4.2. CME RNA-seq enriched genes divided into 4 functional/structural groups.**

| Membrane trafficking/endosomes/lysosome/recycling | Adaptors and core proteins | Plasma membrane scaffolding/cytoskeleton | E3 ubiquitin-protein ligase |
|---|---|---|---|
| Pik3c2a | Dab2 | Snx9 | Cblb |
| Rabep1 | Epn2 | Fnbp1 | Mdm2 |
| Agfg1 | Dvl2 | Pacsin2 | Smurf2 |
| Areg | Stam | Chmp1b | Nedd4l |
| Rab22a | Clta | Hip1 | Rnf41 |
| Pikfyve | Stam2 | Rab11fip3 | |
| Hbegf | Cyth2 | Pard6b | |
| Hspa1b | Ereg | | |
| Chmp2b | Eps15 | | |
| Rab11fip1 | Hspa8 | | |
| Rab5a | Arap2 | | |
| Snx18 | Fcho2 | | |
| Hgs | Cltc | | |
| Rab11a | Smap1 | | |
| Vps37c | Arrb2 | | |
| Rab5b | Arf6 | | |
| Vamp3 | Arfgap1 | | |
| Tor1a | Ldlrap1 | | |
| Scarb2 | Asap2 | | |
| Rab31 | Necap1 | | |
| Vps37b | Ap2a1 | | |
| Eea1 | Il2rb | | |
| F2r | Snx9 | | |

In addition, several genes were annotated on KEGG endocytosis signalling pathway map for quick visual and identification of enriched genes in treated cluster (Figure 3.4.10).
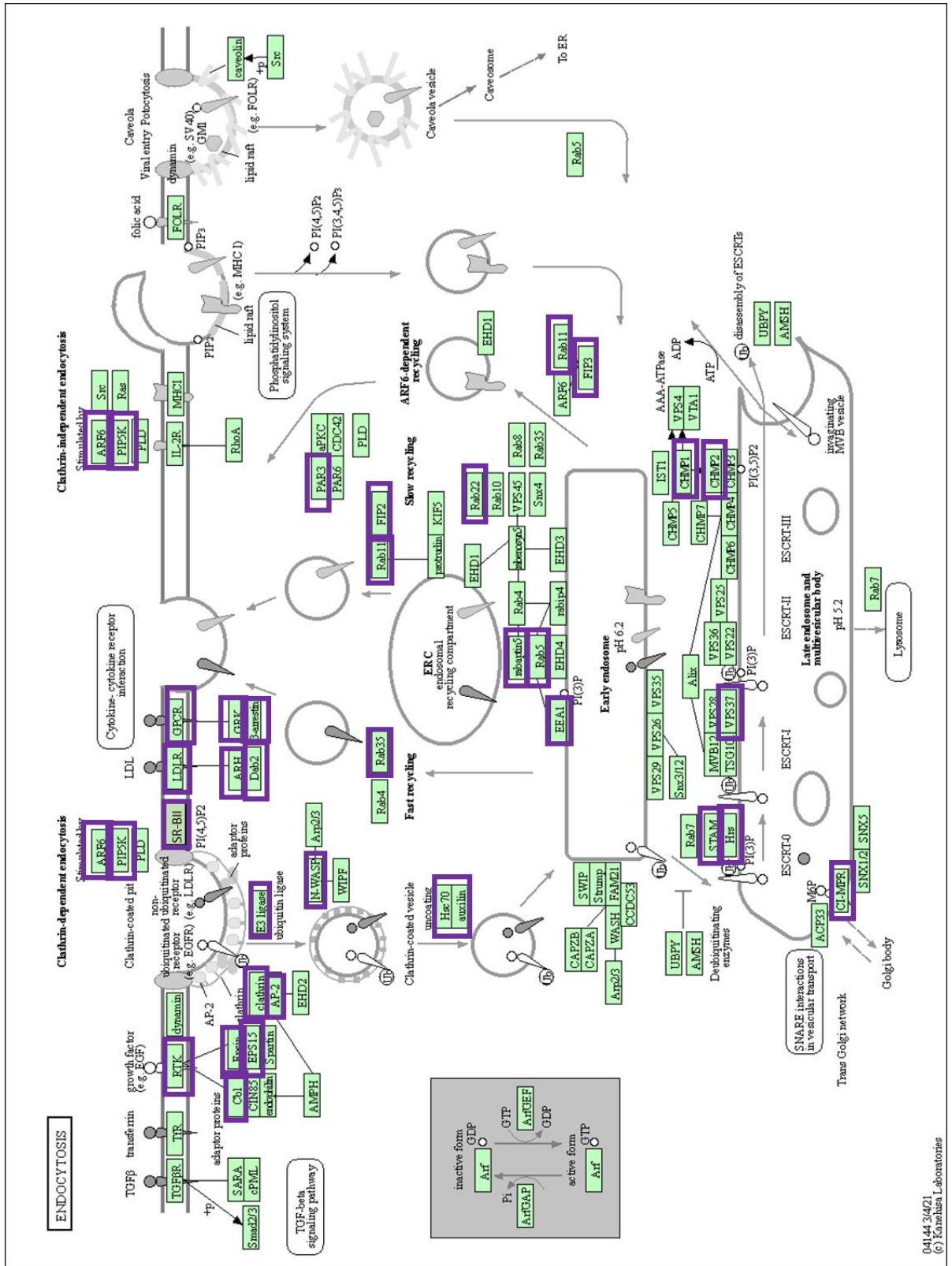
**Figure 3.4.10. Visualisation on KEGG endocytic signalling pathway map.**
*Genes/proteins and their exact impact/effect in endocytosis. Significantly upregulated*

genes found by MSigDB search engine were annotated and highlighted. Figure adapted and modified from (https://www.genome.jp/kegg-bin/show_pathway?mmu04144).

### 3.4.4 NewTagI-II plasma receptor counterpart.

RNA-seq of *Gaa -/-* C2C12 cells following by the gene enrichment analysis unveiled 4 potential plasma membrane receptor candidates to respond to an activation by NewTagI-II motif. Each of the contenders: CXCR4, LDLR, PDGFRa and SR-BII are well known to enter, while stimulated by their ligands, into endocytosis through clathrin-mediated route and potential lysosomal targeting (DeNies et al., 2019; Eckhardt et al., 2006; Goldstein & Brown, 2009; Rakhit et al., 2000).

### 3.4.4.1 Using specific inhibitors to elucidate a potential NewTagI-II plasma receptor interaction.

To determine which one of the aforementioned candidates was the most promising entry point for NewTagI-II protein, an experiment using specific antagonist for each of the chosen receptors was designed. We speculated that by application of a specific receptor antagonist we would block a potential interaction with the receptor and NewTagI-II protein, and therefore block the activation of CME and AP2-P formation.

*Gaa -/-*C2C12cells were differentiated into myofibers and then exposed to NewTagI-II-Myc-Hys peptide in the presence of one of the specific receptor antagonists, experimental conditions and samples are listed in Table 3.4.3. In particular, 1µM BLT-1 inhibitor was used to block SR-BII receptor (Dockendorff et al., 2015), while LDLR was blocked through the treatment with 1µM Receptor associated protein (RAP) (Mokuno et al., 1994), CXCR4 was blocked by specific inhibitor AMD3100 at 5µg/ml (Hatse et al., 2002), and 3µM STI-571 inhibited PDGFRa (Kilic et al., 2000).

**Table 3.4.3. Pharmacological inhibitors.**

| Condition | Inhibitor (final concentration) | NewTagI-II Myc His protein |
|---|---|---|
| UT-control | N/A | N/A |
| SR-BII | 1µM BLT-1 | 80µl |
| LDLR | 1µM RAP | 80µl |
| CXCR4 | 5µg/ml AMD3100 | 80µl |
| PDGFRa | 3µM STI-571 | 80µl |
| Treated -control | N/A | 80µl |

*Gaa -/-* C2C12 cells either untreated or treated with NewTagI-II-Myc-Hys in the presence or absence of a selective inhibitor were collected after 5 min of exposure to NewTagI-II-Myc-Hys and lysed for immunoblotting with anti-AP2-P antibody.

Western blotting membrane revealed that only myofibers treated with SR-BII-inhibitor BLT-1 showed substantial reduction in AP2-P signal compared to treated control (Figure 3.4.11 A). Stripped and reblotted membrane with anti-AP2 antibodies showed no significant difference between any conditions in total signal from AP2 protein. Additional quantification of western blot signal from AP2-P normalised by housekeeping protein GAPDH further supported the observation that level of AP2-P signal was substantially lessened in samples treated with SR-BII receptor's inhibitor (Figure 3.4.11 B).

**A**

| UT | SR-BII | LDLR | CXCR4 | PDGFRa | NewTagI-II Myc His | | |
|----|--------|------|-------|--------|---------------------|---|---|

51 AP2-P
51 AP2
37 GAPDH

**B**
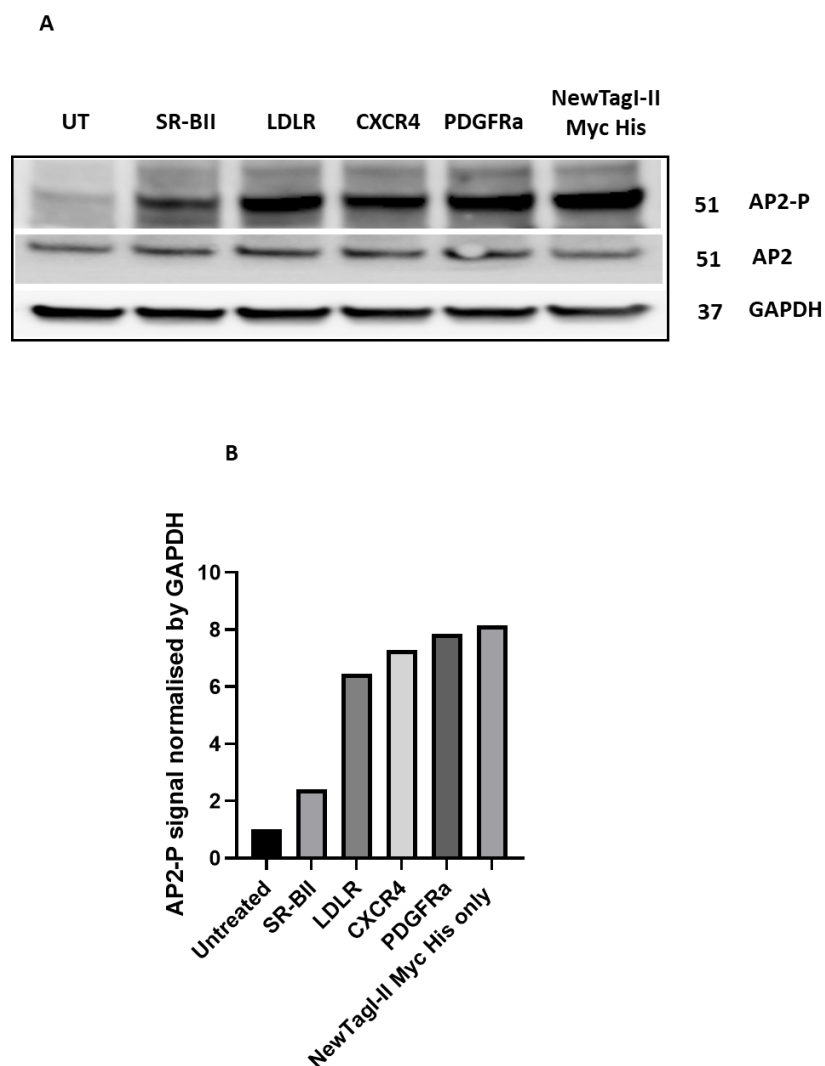
*Figure 3.4.11. Identification of NewTagI-II plasma membrane counterpart receptor. **A)** Western blot showing various levels of AP2-P protein in response to pharmacological inhibitors compared to untreated samples, with particular decreased AP2-P signal in sample treated with SR-BII specific inhibitor. **B)** Quantification of AP2-P bands signal for each condition and normalisation by quantified signal of GAPDH using ImageJ software. Western blotting and its quantification revealed that cells treated with BLT-1 a SR-BII inhibitor and stimulated with NewTagI-II-Myc-His protein, showed fewest signal from AP2-P compared to treated control cells exposed to NewTagI-II Myc His protein only (n=1).*

Based on these preliminary findings we hypothesised that NewTagI-II motif can potentially bind to SR-BII receptor. Although these data would need to be confirmed, it shed light on a new potential entry mechanism into the cells that be used for the delivery of tagged enzymes into lysosomes through gene therapy treatment.

### 3.4.5 Discussion chapter

In study conducted by Gao *et al.* (Gao et al., 2014) the muscle homing peptide (NewTag) showed significant ability to target muscle tissue, and yet in our *in vivo* study addition of NewTag did not improve rhGAA delivery over native GAA molecule. The fundamental question to understand *in vivo* results and to improve GT for PD, was to understand why NewTag did not exhibit its expected action upon skeletal/cardiac muscles as previously reported.

The mechanism of entry of NewTag was established as clathrin-mediated endocytosis (CME) (Figure 3.4.5). C2C12 cells treated with NewTag motif (+/- M6P inhibitor), showed activated CME, however in presence of M6P inhibitor the NewTagI-IIGAA protein did not trigger the CME. Only withdrawal of M6P allowed for CME activation. Unlike the data generated in the Maga study, the NewTag effect of muscle targeting is lost upon combining it with GAA protein. This may suggest that the configuration/design of NewTag with GAA protein is unfavourable and need to be ameliorated for the NewTag motif to deliver its muscle homing targeting.

Following the investigation of a plausible mechanism of NewTag cellular internalisation, the question to answer was to determine the NewTag plasma counterpart/s. There are many ways to exploring direct ligand-receptor interaction i.e., RadioReceptor Assay (RRA) (Ferkany, 1987), crosslinking (Becker & Naider, 2015), and probably the most elaborated and indirect method is global RNA-seq analysis. This method is indirect as it does not detect proteins interaction - ligand-receptor but instead scans transcriptional responses to stimulation caused by effector i.e., ligand NewTag. We decided to use RNA-seq as this method would give use an overview or a snapshot of entire gene expression/RNA production in response to the environmental stimuli/factors. From stimulated C2C12 with NewTag motif the gene set enrichment analysis indicated that 62

genes involved in different aspects of CME are significantly upregulated with 4 genes being CME associated plasma membrane receptors capable of endosomal/lysosomal targeting.

Further experimental analysis in *Gaa -/-* C2C12 cells with selective inhibitors for each individual plasma receptors showed that the only blocking of SR-BII receptor resulted with substantial reduction of the intracellular signalling activated by the treatment with NewTagI-II protein. We therefore hypothesised that SR-BII might be a potential plasma membrane counterpart of NewTagI-II motif.

However, although SR-BII seemed the most probable candidate to interact with NewTagI-II motif, further studies are needed to fully elucidate this claim. Possibly utilising different sets of selective inhibitors would further prove or reject NewTagI-II – SR-BII binding tandem. Furthermore, a greater N number of experimental repetitions would certainly strengthen our hypothesis, as I was only able to conduct n=1 of inhibition experimentation. Nevertheless, due to a timeline restriction on my study I am unable to further explore potential NewTagI-II – SR-BII claim. Moreover, gene set enrichment for downregulated genes was also collected from RNA-seq yet was not analysed for any possible CME genes related. The downregulated genes might had also shed some light on explored areas of CME receptors and/or downstream signalling pathways activated by NewTagI-II modulation. However due to time limitation this analysis would need to be, if any, carried out in the future.

# Chapter 4   Discussion

## 4.1. Human GAA recombinant protein- from transcription to secretion

Currently the gold standard of care for Pompe patients is enzymatic replacement therapy, which alleviates symptoms and allows patients to sustain their existence (Platt et al., 2018). However, ERT has many drawbacks, which include a constant need of hospitalising patients to administer intravenously recombinant GAA protein, an immune response against exogenous rhGAA damping down the effect of ERT and finally, due to a poor expression of CI-M6PR on cell surface, in particular in muscle tissues, a poor cross-correction of the disease in question (Koeberl et al., 2011a) especially in skeletal tissues. There is also a limitation in addressing the CNS component of the disease due to the inability of ERT to cross the BBB. Gene therapy is proposed as an alternative treatment to ERT by providing an endogenous source of GAA proteins to patient's body (Falk et al., 2015; van Til et al., 2010). Current animal studies with intravenous injections of AAV, carrying GAA transgene cassette, targeting liver proved high bioavailability of produced and secreted endogenous GAA protein alleviating PD symptoms. However, AAV is not incorporated into host DNA, but it is limited to episomal DNA only and the therapeutic benefits/amount of produced GAA protein can be diluted down overtime with cell division in a growing organ. Additional boosts with AAV injection suffer from potentially intensive humoral immunogenicity greatly lessening gene therapy outcomes. In addition, GAA recombinant proteins produced by a liver do not possess the ability to cross the BBB limiting the therapeutic outcomes in the CNS (Falk et al., 2015) which has now been shown to greatly contribute to the respiratory pathology observed in Pompe disease (DeRuisseau et al., 2009; Korlimarla et al., 2019). An alternative approach, and one less well explored in gene therapy delivery for PD, is the use of LV vector mediated correction of autologous HSCs. Although LVs incorporating transgenes into the host genome might raise safety concerns in terms of insertional mutagenesis, the use of self-inactivating lentiviral vectors, relatively low vector copy number in the host cells and

internal human promoters have proven to overcome this risk which was seen in historic gammaretroviral vector trials (Baum et al., 2006; Woods et al., 2003). Current studies by Stok *et al.* and van Til *et al.* in a murine model of PD using HSCs and LV vector to transfer GAA transgene cassette proved high bioavailability of endogenous GAA recombinant protein and systemic cross correction. However, in both studies the spleen focus-forming virus (SFFV) promoter was used to drive GAA transgene expression (Stok et al., 2020; van Til et al., 2010). This type of promoter indeed provided a very powerful element to stimulate gene expression, however it was also reported in many pre-clinical studies that it can activate nearby genes in the host genome, including potential oncogenes (Ruscetti et al., 1990; Wolff & Ruscetti, 1988). The SFFV promoter provides great transgene expression, enhancing gene therapy outcomes, nevertheless the risk of mutagenesis is too great and therefore the application of this approach in clinical setting therapies is not suitable.

I explored the possible use of lentiviral vector mediated transduction of HSCs to transfer hGAA transgene equipped with many N-terminal enhancement structures to provide a constant and renewable source of recombinant GAA able to target even tissues generally refractory to ERT by using CI-M6PR independent entry.

Each gene is expressed under the strict control of a specific promoter. Gene expression depends on stress conditions, cellular interactions (Sonawane et al., 2017) and some of the genes being preferentially expressed in specific type of tissues while others are ubiquitously expressed (Montiel-Equihua et al., 2012). I compared GAA expression under two different regulatory sequences: PGK and human β globin LCR-EFS 1α (βLCR-EFS 1α), the former of which has been used in clinical trials, so that any findings can be translated in clinical settings without compromising on patient's safety and wellbeing. Regardless of the cell line used, either K562 (erythroid-like cells) or HEK 293T (kidney cells), *GAA* transgene showed greater expression under the regulation of βLCR-EFS 1α compared to PGK, on average βLCR-EFS 1α showed 50% higher GAA enzymatic activity than PGK (Figure 3.2.3). We hypothesised that while the GAA designs were identical in both cases and only promoters were different, higher GAA enzymatic activity must have been powered by gene promoter activity only. As proven in previous study (Montiel-

Equihua et al., 2012), LCR boosted the EFS expression of controlled genes especially in the erythroid lineage where the β-globin locus control region is active. βLCR-EFS 1α promoter promised a higher gene expression and therefore more optimal gene therapy outcomes.

However, an efficient gene promoter is only one side of the successful story, the other side belongs to a protein itself. The biosynthesis of acid alpha glucosidase is very complicated process requiring several posttranslational modifications including cleavages and glycosylation to finally arrive to be an active form in lysosomal lumens. Only about 10% of the endogenous form of GAA protein leaves its cellular place of biosynthesis as GAA enzyme is not a secretory protein (Martin-Touaux et al., 2002; Moreland et al., 2005). Therefore, to improve its secretion and suit gene therapy's purposes, I manipulated the N-terminus of GAA protein structure to preserve its enzymatic activity and at the same time to incorporate a sequence providing adequate secretion level into outer cellular space and at the same time improving GAA uptake by target cells.

In order to preserve any added modification, I intended to incorporate into the N-terminal of the enzyme in question, a portion of at least 45 amino acids needed to be removed containing a proteolytic site, otherwise any modification added upstream of GAA protein would have been lost in the cleavage event. Based on preclinical study (Maga et al., 2013), which tested several truncated version of GAA protein, I initially tested in my settings three options: a truncation of GAA amino acid structure from 1 to 69 aa, from 1 to 49 or 1 to 54 aa, leaving GAA starting from 70th or 50th/55th amino acid, respectively. The truncation from 1 to 69 aa (GAA70) showed better enzymatic activity compared to the other forms so GAA70 was furtherly employed in the study. Huang *et al.*, Klein *et al.* and Chen *et al.* are only three out of many studies exploring the stability and activity of proteins by using short protein linkers to provide needed support or flexibility to the peptide and its N or C –terminal modifications (Chen et al., 2013; Z. Huang et al., 2016; Klein et al., 2014). In the light of aforementioned studies to improve GAA enzymatic activity and improve protein-protein interaction (NewTagI-IIGAA with its plasma counterpart) , protein linkers were incorporated between the amino acid 70th of GAA and the upstream N-terminal modifications (Chen et al., 2013). I tested two

different protein linkers: one containing three neutral glycine residues and one consisting of 13 amino acids of various polarity. As data showed (Figure 3.1.7), enzymatic activity diminished with the elongation of the linker. The 13aa caused the highest decrease in GAA activity compared to no-protein linker between N-terminus and GAA70, both in cell pellets and medium. The interaction between amino acids in altered recombinant GAA proteins remain an unanswered question and conceivably further study needs to be undertaken through understanding of the enzyme 3D structure and its interlinkage. Despite the inability to improve a catabolic activity of the altered recombinant GAA, I successfully manipulated the N-terminus of GAA protein truncating it to allow an insertion of additional structures responsible for a GAA protein secretion and subsequent uptake, while preserving enzymatic activity.

GAA protein is a non-secretory protein and almost entirely remains intracellularly. After GAA biosynthesis its signal peptide is recognised by a signal peptidase and cleaved to create the complete protein that can be trafficked through the Golgi network targeting lysosomes; only about 10% of GAA is being secreted from the cell via secretory pathways (Gritti, 2011). There are many naturally occurring secretion peptides including alkaline phosphatase or insulin growth factor recognition signal (Pera et al., 2001). The native form of GAA protein does not possess a secretory signal peptide which could potentially jeopardise gene therapy and the mechanism of cross correction (Roig-Zamboni et al., 2017). In order to overcome this hurdle and enforce the secretion of transgenic GAA, I have incorporated a secretory signal into the N-terminus replacing a part of the deleted native amino acids.

I tested different secretion peptides that allow the enzyme safe and abundant departure from the factory cells into the cell medium/ outer cellular space. One consisted of the insulin growth factor II signal peptide (Pera et al., 2001) that was widely study and used. Analysis by Western blotting of the content of GAA protein in cell pellets and in medium of gene modified cells showed that IGF II signal peptide has proven its ability to improve GAA protein secretion; however, a great amount of GAA protein was still present in cell pellets compared to unmanipulated cells. This motivated research for an alternative secretion peptide that would allow higher rate of secretion of GAA protein. Silico secretion peptide (SSP) is a synthetic peptide identified by analysing 50 different

naturally occurring secretion peptides and their aa structures. Hidden Markov model algorithm (Eddy, 1996) used in this study compared all explored sequences and as a result presented with 21 most common aa in all 50 secretion peptides, giving rise to SSP with aa structure as follows MWWRLWWLLLLLLLLLWPMVWA (Guler-Gane et al., 2016). Replacing the IGF II signal peptide with SSP gave significantly improved results. Similar to Guler-Gane *et al.* findings, I showed 2.5-fold increase in GAA protein secretion when SSP was cloned instead of IGF II signal peptide into equivalent rhGAA constructs (Figure 3.1.7). In addition, the amount of GAA accumulated in cells was reduced and comparable with untreated K562 cells. Furthermore, to evaluate GAA gene expression under SSP/IGF II secretion peptides I conducted mRNA study, which revealed that neither SSP nor IGF II bearing GAA constructs differ from native GAA gene expression level further supporting SSP use as a more robust secretion peptide that does not interfere with gene expression.

Posttranslational modification (PTM), following transcription, is a critical step in the maturation process of newly formed functional proteins relying on chemical alterations, regulating protein stability and function. Wang *et al.* and Hannoun *et al.* studied in depth PTM showing that any deviation from cellular pathways or intrusion into biochemical processes may lead to a partial/full loss of protein biological function (Hannoun et al., 2010; Wang et al., 2014). Following these studies, I decided to analyse posttranslational modifications of SSP GAA recombinant protein aiming to investigate N-linked glycosylation patterns, in particular the attachment of high mannose oligosaccharides on SSP/IGF/Native GAA proteins. High mannose oligosaccharides when phosphorylated in a position 6 on the sugar's complex structure can be recognised by M6PR present on cell surface and utilised as an endocytic pathway activator, therefore it has a relevant implication in the efficacy of cross correction. Following Bozon *et al.* I used endoglycosidases to digest different GAA recombinant proteins secreted by transduced K562 cells into medium. WB analysis of the digested products showed no differences in bands sizes between manipulated GAA protein and native GAA (Figure 3.2.2) (Bozon et al., 1986). We hypothesised based on all experiments including SSP GAA that SSP,

despite its powerful secretory potential, neither affects the glycosylation process nor the correct protein folding.

The SSP GAA recombinant protein design proved to be enzymatically active, with mRNA level similar to the native GAA and 2.5-fold higher secreted than IGF II GAA equivalent constructs, proving to be a very good platform for cross correction in gene therapy in PD. However, a total spectrum of PTM of SSP GAA protein would be required to full assess methylation, O-glycosylation, phosphorylation to understand comprehensive range of SSP motif impact on GAA protein, if any.

## 4.2 Cross correction and alleviation of glycogen burden *in vitro*

Koeberl *et al.* in his work on PD studied CI-M6PR and its poor expression on muscle fibers (Koeberl et al., 2011a) that in turn significantly decreased ERT outcomes in management of the disease in question (Koeberl et al., 2011a; Marques & Saftig, 2019). To date CI-M6PR is the only recognised and commercially utilised receptor that allows glycosylation-dependent (through M6P) lysosomal targeting as well as an alternative glycosylation-independent lysosomal transport (GILT) (LeBowitz et al., 2004; Maga et al., 2013). Maga *et al.* investigated the use of IGF II part of CI-M6PR and showed reduction in glycogen accumulation in treated mice. However, utilising different part of CI-M6PR will not make up for a low total receptor expression on cell surface seen in PD patients. Therefore, I decided to investigate alternative pathways to CI-M6PR by using a muscle homing peptide (MHP), already described in literature by Gao *et al.* to drive morpholino into muscle tissues in dystrophic mice (Gao et al., 2014). Although the MHP mechanism of action remains unknown, following Gao *et al.* I incorporated MHP structure, called NewTag in this study, which is composed of 12 amino acids (RRQPPRSISSHP) into GAA construct. I tested several designs with single NewTag domain (Newtag I) or two NewTags structures in tandem (NewTag I-II), all of which were human codon optimised. In confocal microscopy experiments using H2K 2B4 myofibers exposed for 24 hrs to NewTag I or NewTagGAA I-II GAA protein coupled with mCherry reporter in the presence or absence of CI-M6PR competitive antagonist, I tested the uptake of GAA recombinant proteins (Figure 3.1.6). NewTag facilitated GAA uptake and lysosomal trafficking of the

recombinant protein, which still occurred in presence of CI-M6PR competitive antagonist, with the tandem design showing the best uptake (Figure 3.1.7). These data indicated that the GAA protein equipped with the new MHP benefits from two distinct pathways of uptake and lysosomal trafficking: one still unknown and muscle specific and the other one still dependent on CI-M6PR. Keeping both options for entering cells can maximise cross correction and future gene therapy outcomes.

Further analyses of rhGAA uptake and its intracellular/lysosomal enzymatic activity including glycogen reduction was conducted in *Gaa -/-* C2C12 cell line, generated for the purpose of this study using the now well established CRISPR/Cas9 gene editing technique. CI-M6PR expression on *Gaa -/-* C2C12 cells surfaces was investigated by western blot and confocal microscopy showing no visual differences with wild type C2C12 cell line and suggesting that gene editing did not alter the receptors' distribution and hence this cell line could be used as *in vitro* model of PD.

Uptake studies using *Gaa -/-* C2C12 cells in a transwell system with transduced K562 cell lines that limited the passage to only solutes showed that GAA activity in *Gaa -/-* C2C12 differed after exposure to rhGAA proteins bearing different designs of N-terminal structures. *Gaa -/-* C2C12 cells exposed to SSP NewTag I-IIGAA recombinant proteins displayed the highest values of GAA activity that was significantly different to its equivalent construct bearing IGF II signal peptide, suggesting the benefit of using SSP and also showing NewTag as a potential alternative to M6P muscle entry point. Although NewTag was described as a muscle specific homing peptide, I could measure only ~ 1% of GAA activity in *Gaa -/-* C2C12 in transwell setting, in presence and absence of M6P inhibitor, compared to wild type myoblast cell line. Perhaps, this might be due either to biochemical interference of GAA processing upon modification or to the inability of NewTag to strongly interact with its cellular counterpart.

To determine a level of aforementioned rhGAA uptake into *Gaa -/-* cell line and its lysosomal targeting, I utilized a protocol of starvation assay developed by Yoshida *et al.* study to quantitate glycogen reduction. This starvation assay employed a glucose and

pyruvate free DMEM and dialysed FBS to stop glycogen formation and remove any glycogen and glucose from the cellular matrix, but not glycogen in lysosomes (Yoshida et al., 2017). In starvation condition, *Gaa -/-* C2C12 exposed to SSP NewTag I-IIGAA showed 70% decrease in glycogen concentration compared to starved untreated *Gaa -/-* C2C12 control cells and *Gaa -/-* C2C12 exposed to IGFII NewTagI-IIGAA (Figure 3.2.8). Notably, these data on the reduction of glycogen in lysosomes were observed in the presence of CI-M6PR antagonist suggesting that SPP NewTag I-IIGAA recombinant protein was successfully taken up by myofibers and subsequently directed into lysosomes, through a yet not known pathway/ mechanism.

The newly designed N-terminal structures of GAA recombinant protein were proved to be able to target the lysosomal compartment and subsequently alleviate a lysosomal glycogen burden in an *in vitro* cell model of PD. This *in vitro* cross correction model utilising a modified rhGAA to rescue GAA enzymatic activity in *Gaa -/-* cell line and successfully decrease its lysosomal glycogen accumulation indicate a good potential to use SSP NewTagI-IIGAA in HSCs gene therapy for PD. However, at this stage it is difficult to clarify if the observed levels of rescued GAA enzyme and reduction of glycogen in *Gaa -/-* cell line are adequate for subsequent clinical translation in PD patients.


### 4.3. Cross correction and alleviation of glycogen burden *in vivo*

To date there are very a few LV gene therapy studies utilising transplant of HSCs for PD. Stock *et al.* showed full correction of PD in both functional and metabolic aspect of the disease in questions in a mouse PD model (Stok et al., 2020) . However, to achieve this significant correction high VCN (around 8 copies of vector per cell) and a viral SF LTR based promoter, to drive high transgene expression, was used, which from clinical point of view is of a particular concern. In a second study, Piras *et al.*, using a native, unmodified structure of hGAA protein showed partial correction of PD in cardiac and skeletal muscles in a mouse model of PD, with relatively low VCN and the use of βLCR-EFS 1α regulatory cassette for transgene expression with a clinically relevant vector design (Piras et al., 2020).

However, both aforementioned studies relied entirely on CI-M6PR to deliver a cross correction. This receptor is not greatly expressed on skeletal muscle and M6P binding motif does not show great binding affinity to its ligand and, as a consequence, gene therapy requires the use of strong promoters that are able to drive very high GAA expression to fully correct the disease (Gary-Bobo et al., 2007; Ghosh et al., 2003; Nykjaer et al., 1998). Therefore, in this study a mammalian promoter construct consisting of EFS1 and β-globin locus control region was used to drive high transgene expression which is enhanced in erythroid lineages. In addition, a strong SSP secretion peptide provided a paramount level of secreted, correctly glycosylated rhGAA protein in the system. Furthermore, a novel muscle homing peptide, already proven *in vivo* to target muscle tissues (Gao et al., 2014; Jirka et al., 2018) was applied at N-terminus of truncated GAA protein to further support internalisation and cross correction. This new SSPNewTagI-II GAA construct was tested *in vivo* in the *Gaa -/-* mouse PD model described in Piras *et al.* (Piras et al., 2020).

Our initial/optimisation study with HSCs transduced with LV SSPNewTagI-II GAA at MOI 150 showed in analysed LC and CFU relatively low VCN varying between 1-2 copies of vectors per cell. Also, vector containing the GAA native only structure, that was used to establish a control base line for correction of PD, showed very low VCN. Sweeney *et al.* studying the impact of lentiviral vector genome size on viral integration and VCN, suggested that a greater size of transgene cassette can negatively impact unpacking viral genome and interfere with reverse transcription of viral material in transduced cells (Sweeney & Vink, 2021). We came to the conclusion that our insert consisting of 2.5kb regulatory cassette prompter (βLCR-EFS 1α) and 3.5kb transgene (recombinant hGAA) increased the size of the viral vector by ~6kb leading to poor lentiviral integration/cellular internalisation causing the low VCN observed (Ghosh et al., 2020; Papayannakos & Daniel, 2013). In order to achieve higher VCN for the benefit of a successful metabolic correction (Piras, Chan, Wantuch et al., 2020; Stok et al., 2020) , we decided to make the insert significantly smaller by exchanging βLCR-EFS 1α for PGK promoter (~500bp). Additional benefit of using PGK promoter in this study was the fact that it is already widely tested *in vitro* and used in clinical studies (Garcia-Perez et al., 2020).

As expected, the transduction of HSCs with LV PGK- SSPNewTagI-II GAA or LV PGK- GAA native at MOI150 revealed that the reduction in vector size by the switching of promoter led to an increased VCN of ~6 copies per cell.

We then compared the cross-correction of *Gaa -/-* HSCs corrected with either LV PGK-SSPNewTagI-II GAA or LV PGK- GAA in a mouse model of PD. As 4 months post-transplant with transduced HSCs is sufficient to correct PD in skeletal muscle (Stok et al., 2020), we decided to set *in vivo* study end point at 3 months post in contrast to our initial plan of 6 months post-transplant.

Initial data on engraftment, using Y chromosome probes showed consistent and stable engraftment (100%) of donors male HSCs into females' recipients. In addition, VCN acquired from PBMC and BM at the end of this study was in line with data obtained from *in vitro* cultures of the HSCs (both LC and CFUs) prior infusion. Also, a panel of cellular markers applied to white blood cells collected from spleens of experimental animals show no significant variation in the numbers of myeloid, T and B cells compared to untreated *Gaa -/-* animals. These findings were of paramount importance as provided evidence that the vectors carried transgene expression were, firstly, still present in transplanted HSC and therefore can support long term GT goals, and secondly, vectors used did not skew/impact upon relative proportions of differentiated blood lineage progeny formed from transduced and transplanted cells compared to untransduced *Gaa -/-* mice. The secretion peptide SSP, as expected, delivered a strikingly high level of rhGAA protein in plasma, measured by GAA activity. Also, both rhGAA constructs, GAA native and SSPNewTagI-II GAA, purified from experimental animals' plasma and treated by endoglycosidase PNGase and Endo H, showed no difference in their N-glycosylation pattern by western blot band sizes analysis.

Functional analysis of rotarod of animals in this cohort showed disappointingly that GT treated animals with both vectors did not WT latency level and being of no significant difference to *Gaa -/-* untreated control animals. However ,interestingly Glc4 level, that is a product of degradation of glycogen and it is one of the biomarkers of PD tested in urine (Kishnani, Steiner, et al., 2006; Tarnopolsky et al., 2016), was nearly significantly lower in SSPNewTagI-IIGAA group (p=0.08) compared to *Gaa -/-* mice and significantly

lower compared to GAA native treated group (p<0.01,+/-SEM). In contrast, there was no statistical significance observed between GAA native treated group vs *Gaa -/-* mice.

Disappointingly, echocardiography data showed no major differences between GT treated cohorts, WT and *Gaa -/-* groups. In contrast, Piras *et al.* (Piras et al., 2020) showed clear separation between *Gaa -/-* and other cohorts of the study using the same PD mouse model. However, on the close inspection Piras *et al.* conducted echocardiography readings on 6-month-old animals in contrast to 3-months-old in this study. We hypothesise that the lack of difference between *Gaa -/-* mice and other groups might be due to fact that cardiac muscle did not accumulate sufficient glycogen related damage yet and cardiac functions in *Gaa -/-* mice were not sufficiently affected to allow a difference to be shown. Unfortunately, no published data/papers exist to date to support or reject this claim and a further extended study would need to be conducted for comparison.

An interesting body of evidence was delivered from homogenised harvested tissues. GAA activity measured in individual sample of homogenise and normalised by a total amount of proteins concentration revealed that both GT groups showed moderate GAA activity in comparison to *Gaa -/-* untreated animals, however this GAA activity was only a very small fraction of the GAA activity shown by samples from WT animals. Some of the GT treated animals/samples showed in diaphragm, lungs and liver, substantial GAA activity; however, it is likely that this is an artefact of residual rhGAA carried from blood cells that was insufficiently flushed out in systemic PBS perfusion, performed at the end point to clear the total GAA activity per given homogenised organ from the contamination of GAA activity of donor blood cells. This conclusion was based in accordance with glycogen measurements on homogenised tissues showing no substantial glycogen reduction in given samples, as discussed below.

Glycogen accumulation is a hallmark of PD and its reduction is crucial to relieve PD symptoms (Lim et al., 2014a). It is clear now that not only skeletal and cardiac muscle accumulation of glycogen contribute to PD symptoms, but a new body of evidence suggests that accumulation of glycogen in CNS, and in particular in brain, might also greatly impact patient's cognitive and psychomotor functions and contributes to the overall of PD symptoms (Favret et al., 2020; Korlimarla et al., 2019). However, it is not

fully known yet how much of lysosomal glycogen accumulation is allowed without causing impairment of lysosomal and/or endosomal structures, trafficking, and general cellular homeostasis (Gungor et al., 2011; Lim et al., 2014a; Peruzzo et al., 2019).

Glycogen concentration was conducted on homogenised tissues previously used to establish GAA activity. Glycogen readouts were measured and normalised by a total protein per each sample. GT treated animals in both groups showed various reduction ranging from 30 to 50% in glycogen concentration compared to *Gaa -/-* untreated control animals. Heart, diaphragm, tibialis anterior and gastrocnemius responded moderate to well to two of the rhGAA protein versions. All tissues showed statistically significant drop in glycogen accumulation in GT treated groups compared to *Gaa -/-* mice (p<0.01, +/-SEM, n=5 SSPNewtagI-IIGAA and n=4 GAA native, n=6 *Gaa -/-*). Most interestingly brain samples, that normally this organ is out of reach for standard ERT treatment due to BBB restrictions, showed also significantly less glycogen detected in GT groups compared to *Gaa -/-* untreated control. This drop in glycogen detected in GT groups suggests that even moderate amount of rhGAA detected in homogenised tissues is capable of alleviating, to a certain extent, glycogen burden. Our hypothesis that insufficient perfusion and blood rhGAA residues contributed to the overall of GAA activity detected in lungs, liver, and diaphragm of some of the GT animals was indirectly proven by the relatively high concentration of glycogen measured in corresponding organs.

Overall, both constructs used in this study were proven to be able to express transgene over a long period of time and not be silenced, and also none of the constructs/vectors skewed blood lineages formed from transduced and transplanted HSCs. In addition, both constructs provided constant and renewable source of functional and correctly glycosylated rhGAA in circulation, that in turn reduced in various levels glycogen accumulation in tested tissues.

This study and employed model of HSC GT answered many valuable questions in regard to the stability and durability of transgene expression in hematopoietic cells, the lack of vector toxicity onto blood lineages distribution, and the protein's engineering design, including their stability and effective secretion power. However, this *in vivo* study, conversely to what seen in the in vitro study, showed no greater potential of the new

form of recombinant GAA to overcome the hurdle of CI-M6PR uptake limitation and increase muscle targeting and rhGAA internalisation. The proposed muscle homing peptide, despite several publication (Gao et al., 2014; Jirka et al., 2018) did not prove to facilitate muscle targeting above and beyond what is observed with GAA native construct. Perhaps small number of animals in each group did not provide sufficient power to differentiate clearly between two tested constructs and *Gaa -/-* animals. Furthermore, it also seems that the design of fusion protein and perhaps its individual components were suboptimally configured and did not allow substantial cross-correction with GAA enzymatic activity/glycogen level comparable to WT animals.

On the other hand, *in vivo* findings seem to confirm our i*n vitro* data from transduced K562 with SSPNewTagI-II GAA showing high levels of secretion and high gene expression, although *Gaa -/-* C2C12 treated with MHP GAA proteins in presence of M6P inhibitor in a transwell system showed only modest increase in GAA activity and glycogen reduction compared to wild type C2C12.

One of the most puzzling questions is related to a high level of NewTagI-IIGAA recombinant protein detected in blood plasma of experimental animals and yet this did not translate into high GAA activity in homogenised tissues. It is known that despite the high dose of rhGAA enzyme particles being administrated as ERT during hospital administration, PD correction in majority individuals is very limited (Do et al., 2019), suggesting perhaps that the *in vivo* system has a saturation point above which there is no more migration of rhGAA through endothelial lining of blood vesicles, and yet with low CI-M6PR expression in muscle tissue the amount of rhGAA getting through does not deliver sufficient reduction in glycogen. Perhaps, regardless of correct glycosylation, the stability of the NewTagI-IIGAA protein is, by some means, affected and less penetration of endothelial lining is allowed. These fundamental questions regarding proteins stability in plasma and penetration potential ought to be a focus point in a future study to fully elucidate this matter.

Another interesting question relates to MHP 'the NewTagI-II' motif itself. It was proven in many studies that this short 12 amino acids fragment increased significantly a muscle targeting and internalisation of its cargo (Gao et al., 2014; Jirka et al., 2018). Further studies to address this question were undertaken.

**4.4 Muscle homing peptide 'NewTag' motif and its potential plasma counterpart**

SSPNewTagI-II GAA recombinant protein *in vitro* study of transduced K562 cell line showed excellent secretion, high GAA enzymatic activity and correct N-linked glycosylation compared to native form of GAA protein. In addition, in the transwell study SSPNewTagI-II GAA showed moderate level of internalisation and reduction of lysosomal glycogen in treated *Gaa -/-* C2C12 cells.

However, despite good VCN in transduced and transplanted HSCs, high level of secreted protein in animals' blood plasma and its correct N-linked glycosylation compared to native GAA construct, the SSPNewTagI-II GAA recombinant protein did not substantially decrease glycogen accumulated in experimental animals' tissues compared to animals treated with GAA native construct.

Determination of NewTag cellular entry point/mechanism was a paramount task in order to put forward possible explanation(s) for an inadequate internalisation and in consequence suboptimal glycogen burden reduction *in vivo.* Further studies using NewTagI-II Myc His motif and *Gaa -/-* C2C12 cells shed more light on the observed NewTag motif limitations to increase cellular internalisation and glycogen level reduction in *Gaa -/-* mice. Applying several entry point/mechanisms inhibitors, we dissected that the cell entry route of NewTag is mediated by a clathrin-related endocytosis (CME) mechanism. Further to this discovery, AP2/AP2-P core CME protein was used as a marker of CME activation (Kaksonen & Roux, 2018), and we hypothesised that NewTag triggering CME through the binding to its yet unknown receptor would stimulate the phosphorylation of AP2 protein. This phosphorylation is crucial for activation of AP2 and in turn the assembly of CME (Kadlecova et al., 2017). Several consecutive experiments exposing *Gaa -/-* C2C12 cells to series of different conditions, including NewTagI-II Myc His protein vs NewTagI-IIGAA enzyme with or without M6P inhibitor, were undertaken to determine AP2-P level in treated cells. Phosphoprotein's half-life is measured in minutes or even seconds (Gelens & Saurin, 2018), therefore to evaluate activation of CME and phosphorylation of AP2 core protein, cells were

harvested after only 5 minutes post-treatment with NewTagI-II Myc His or NewTagI-IIGAA protein, and their lysate assed by Western blot.

NewTagI-II Myc His protein clearly showed the activation of CME as detected by the significant increase of AP2-P level compared to untreated samples, regardless of the presence or absence of M6P inhibitor. In contrast, NewTagI-IIGAA protein in presence of M6P inhibitor did not trigger CME as showed by the level of AP2-P signal comparable to untreated cell samples. However, the withdrawal of M6P inhibitor from cell culture medium, generated the increase in AP2-P signal in cells treated with NewTagI-IIGAA protein compared to untreated sample. This series of experiments demonstrated a potential of NewTag motif alone to induce AP2-P formation indicating its mechanisms of entry. However, combining NewTag with GAA protein seems to be undermining potential of NewTag's CME entry point. Further analysis of the structure of NewTagI-IIGAA protein would be necessary to determine if/how the NewTag motif is disturbed by its conjugated 'big brother' GAA enzyme.

From the gathered evidence the most probable explanation to why animals exposed to NewTagI-IIGAA protein showed similar level in reduction of glycogen to the one exposed to GAA native protein is that internalisation of NewTagI-IIGAA recombinant was primarily induced by M6P structures on GAA protein and not NewTag motif. Therefore, high level of circulating NewTagI-IIGAA recombinant protein was an insufficient factor to correct PD in mice because still limited to the poor expression of CI-M6PR on muscle cells and its low affinity for M6P ligand (Gary-Bobo et al., 2007).

On a first glimpse this finding stands in contrast to *in vitro* findings were NewTagI-IIGAA showed moderately higher GAA activity and glycogen reduction in *Gaa -/-* C2C12 compared to GAA native treated cells. However, the transwell *in vitro* experiments were set up for 24hrs and therefore in these longer duration experiments NewTag might exhibit additional support to GAA uptake not seen in a very short 5 minutes exposure.

However, the additive effect to cellular uptake of NewTag motif was somehow mitigated in *in vivo* PD study; but the value of this motif should not be underestimated in perhaps different design and protein's configuration with its cargo partner. Therefore, studies to elucidate the NewTag's plasma membrane counterpart was undertaken in a serious of

experiments leading to total RNA sequencing and gene enrichment analysis of upregulated genes in NewTagI-II Myc His treated vs untreated *Gaa -/-* C2C12 cells.

From the analysis of total RNA sequencing, data suggested that genes responsible for several proteins involved in CME, including heavy and light chain of clathrin and AP2 core protein, are significantly upregulated in NewTagI-II Myc His treated compared to untreated cells. This discovery confirmed our previous findings from *in vitro* analysis with endocytosis inhibitors proving that MHP – 'NewTag' is more than capable of activating CME and cellular internalisation. Our findings support earlier published study by Gao *et al.* (Gao et al., 2014) in dystrophin restoration by muscle homing peptide-morpholino conjugated probe proving capacity of NewTag to be an effective drug delivery peptide.

Furthermore, within our RNA seq and gene enrichment analysis, several CME plasma membrane receptors were revealed, giving rise to a final set of experiments aiming at identification of the potential NewTag plasma counterpart. Four receptors were interrogated including Chemokine receptor 4 (CXCR4) (DeNies et al., 2019), Low density lipoprotein receptor (LDLR) (Goldstein & Brown, 2009), Platelet derived growth factor receptor α (PDGFRa) (Rakhit et al., 2000) and Scavenger receptor BII (SR-BII) (Eckhardt et al., 2006). SR-BII came out to be the most enigmatic find, as in fact the gene enrichment analysis showed only SR-BI gene to be statistically upregulated, however this receptor is not a clathrin-dependent endocytic receptor, yet still an endocytic receptor (Zani et al., 2015). Following a thorough literature review, it came to light, that SR-BII receptor is encoded by an alternatively spliced mRNA of the SR-BI gene and shares with SR-BI identical extracellular domain structure, with a difference in the cytoplasmic carboxyl-tail only. More importantly SR-BII is located to clathrin-coated pits on the cell membrane and triggers CME (Eckhardt et al., 2006). By using specific receptor inhibitors for those four aforementioned receptors, we found that blocking SR-BII reduced AP2-P formation in *Gaa -/-* myofibers cells treated with NewTagI-II Myc His protein (Figure 3.4.11). This experiment suggested the SR-BII is potentially responsible for the internalisation of NewTagI-II motif. However, further studies are required to fully elucidate this claim. Perhaps, different time points for total RNA collection from treated cells would be to validate or reject endocytosis genes enrichment in treated cluster. In

addition, further study on SR-BII i.e., kinetics and free energy release from ligand-receptor interaction; generating a SR-BII knockout C2C12 cell line model to determine activation of CME while treated with NewTagI-II motif; and finally, examination of different specific antagonists of proposed candidate receptor would also add to a validation/rejection of the final claim regarding NewTagI-II motif and its interaction with the proposed plasma membrane receptor.

Unfortunately, the proposed additional steps to validate the findings from this chapter would greatly exceed a timeframe of this PhD project and should be continue in different circumstances.

# 5.Bibliography

Abramowicz, A., & Gos, M. (2019). Correction to: Splicing mutations in human genetic disorders: examples, detection, and confirmation. *J Appl Genet, 60*(2), 231.

Adeva-Andany, M. M., et al. (2016). Glycogen metabolism in humans. *Bba Clinical, 5*, 85-100.

Agbandje-McKenna, M., & Kuhn, R. (2011). Current opinion in virology: structural virology. *Curr Opin Virol, 1*(2), 81-83.

Aguilar, R. C., & Wendland, B. (2005). Endocytosis of membrane receptors: Two pathways are better than one. *Proceedings of the National Academy of Sciences of the United States of America, 102*(8), 2679-2680.

Aiuti, A., et al. (2013). Lentiviral hematopoietic stem cell gene therapy in patients with Wiskott-Aldrich syndrome. *Science, 341*(6148), 1233151.

Al Khallaf, H., et al. (2013). CRIM-negative Pompe disease patients with satisfactory clinical outcomes on enzyme replacement therapy. *Molecular Genetics and Metabolism, 108*(2), S18-S18.

Alberts, B., et al. (2015). Molecular Biology of the Cell, Sixth Edition. *Molecular Biology of the Cell, Sixth Edition*, 1-1342.

Amalfitano, A., et al. (1999). Systemic correction of the muscle disorder glycogen storage disease type II after hepatic targeting of a modified adenovirus vector encoding human acid-alpha-glucosidase. *Proc Natl Acad Sci U S A, 96*(16), 8861-8866.

Amarante-Mendes, G. P., et al. (2018). Pattern Recognition Receptors and the Host Cell Death Molecular Machinery. *Frontiers in Immunology, 9*.

Andersson, L. C., et al. (1979). K562--a human erythroleukemic cell line. *Int J Cancer, 23*(2), 143-147.

Angelini, C., & Engel, A. G. (1973). Subcellular distribution of acid and neutral alpha-glucosidases in normal, acid maltase deficient, and myophosphorylase deficient human skeletal muscle. *Arch Biochem Biophys, 156*(1), 350-355.

Anjum, A., et al. (2016). Identification of Differentially Expressed Genes in RNA-seq Data of Arabidopsis thaliana: A Compound Distribution Approach. *Journal of Computational Biology, 23*(4), 239-247.

Anna, A., & Monika, G. (2018). Splicing mutations in human genetic disorders: examples, detection, and confirmation. *J Appl Genet, 59*(3), 253-268.

Aronovich, E. L., et al. (2017). Prolonged Expression of Secreted Enzymes in Dogs After Liver-Directed Delivery of Sleeping Beauty Transposons: Implications for Non-Viral Gene Therapy of Systemic Disease. *Hum Gene Ther, 28*(7), 551-564.

Asheuer, M., et al. (2004). Human CD34+ cells differentiate into microglia and express recombinant therapeutic protein. *Proc Natl Acad Sci U S A, 101*(10), 3557-3562.

Atchison, R. W., et al. (1965). Adenovirus-Associated Defective Virus Particles. *Science, 149*(3685), 754-756.

Aung-Htut, M. T., et al. (2020). Splice modulating antisense oligonucleotides restore some acid-alpha-glucosidase activity in cells derived from patients with late-onset Pompe disease. *Sci Rep, 10*(1), 6702.

Azzouz, M., & Mazarakis, N. (2004). Non-primate EIAV-based lentiviral vectors as gene delivery system for motor neuron diseases. *Curr Gene Ther, 4*(3), 277-286.

Baeza-Centurion, P., et al. (2020). Mutations primarily alter the inclusion of alternatively spliced exons. *Elife, 9*.

Baik, A. D., et al. (2021). Cell type-selective targeted delivery of a recombinant lysosomal enzyme for enzyme therapies. *Mol Ther, 29*(12), 3512-3524.

Ball, S., et al. (2011). The evolution of glycogen and starch metabolism in eukaryotes gives molecular clues to understand the establishment of plastid endosymbiosis. *Journal of Experimental Botany, 62*(6), 1775-1801.

Ballabio, A., & Gieselmann, V. (2009). Lysosomal disorders: from storage to cellular damage. *Biochim Biophys Acta, 1793*(4), 684-696.

Balliu, B., et al. (2019). Genetic regulation of gene expression and splicing during a 10-year period of human aging. *Genome Biology, 20*(1).

Baralle, D., & Buratti, E. (2017). RNA splicing in human disease and in the clinic. *Clin Sci (Lond), 131*(5), 355-368.

Barry, M. A., et al. (1996). Toward cell-targeting gene therapy vectors: selection of cell-binding peptides from random peptide-presenting phage libraries. *Nat Med, 2*(3), 299-305.

Barton, N. W., et al. (1991). Replacement therapy for inherited enzyme deficiency--macrophage-targeted glucocerebrosidase for Gaucher's disease. *N Engl J Med, 324*(21), 1464-1470.

Baum, C., et al. (2006). Mutagenesis and oncogenesis by chromosomal insertion of gene transfer vectors. *Hum Gene Ther, 17*(3), 253-263. doi:10.1089/hum.2006.17.253

Beck, M. (2010a). Emerging drugs for lysosomal storage diseases. *Expert Opin Emerg Drugs, 15*(3), 495-507.

Beck, M. (2010b). Therapy for lysosomal storage disorders. *IUBMB Life, 62*(1), 33-40.

Becker, J. M., & Naider, F. (2015). Cross-linking Strategies to Study Peptide Ligand-Receptor Interactions. *Membrane Proteins - Production and Functional Characterization, 556*, 527-547.

Beratis, N. G., et al. (1978). Characterization of the molecular defect in infantile and adult acid alpha-glucosidase deficiency fibroblasts. *J Clin Invest, 62*(6), 1264-1274.

Berro, J. (2019). Uncovering the Mechanisms of Clathrin-Mediated Endocytosis using Quantitative Biology Approaches. *Biophysical Journal, 116*(3), 150a-150a.

Biffi, A. (2017). Hematopoietic Stem Cell Gene Therapy for Storage Disease: Current and New Indications. *Mol Ther, 25*(5), 1155-1162.

Biffi, A., et al. (2006). Gene therapy of metachromatic leukodystrophy reverses neurological damage and deficits in mice. *J Clin Invest, 116*(11), 3070-3082.

Biffi, A., et al. (2013). Lentiviral hematopoietic stem cell gene therapy benefits metachromatic leukodystrophy. *Science, 341*(6148), 1233158.

Blaese, R. M., et al. (1995). T lymphocyte-directed gene therapy for ADA- SCID: initial trial results after 4 years. *Science, 270*(5235), 475-480.

Boentert, M., et al. (2016). Pattern and prognostic value of cardiac involvement in patients with late-onset pompe disease: a comprehensive cardiovascular magnetic resonance approach. *J Cardiovasc Magn Reson, 18*(1), 91.

Bowen, D. G., et al. (2002). Cytokine-dependent bystander hepatitis due to intrahepatic murine CD8(+) T-cell activation by bone marrow-derived cells. *Gastroenterology, 123*(4), 1252-1264.

Boyce, W. T., et al. (2020). Genes and environments, development and time. *Proceedings of the National Academy of Sciences of the United States of America, 117*(38), 23235-23241.

Bozon, D., et al. (1986). Characterization of cellular oligosaccharides from normal and cystic fibrotic fibroblasts using sequential endoglycosidase digestions. *Arch Biochem Biophys, 249*(2), 546-556.

Braulke, T., & Bonifacino, J. S. (2009). Sorting of lysosomal proteins. *Biochim Biophys Acta, 1793*(4), 605-614.

Brooks, D. A., et al. (2006). Stop-codon read-through for patients affected by a lysosomal storage disorder. *Trends Mol Med, 12*(8), 367-373.

Bryder, D., et al. (2006). Hematopoietic stem cells: the paradigmatic tissue-specific stem cell. *Am J Pathol, 169*(2), 338-346.

Bryois, J., et al. (2017). Time-dependent genetic effects on gene expression implicate aging processes. *Genome Research, 27*(4), 545-552.

Burke, L. M., et al. (2017). Postexercise muscle glycogen resynthesis in humans. *Journal of Applied Physiology, 122*(5), 1055-1067.

Burton, B. K., et al. (2020). Newborn Screening for Pompe Disease in Illinois: Experience with 684,290 Infants. *Int J Neonatal Screen, 6*(1), 4.

Bushman, F. D. (2007). Retroviral integration and human gene therapy. *J Clin Invest, 117*(8), 2083-2086.

Byrne, B. J., et al. (2019). Pompe disease gene therapy: neural manifestations require consideration of CNS directed therapy. *Ann Transl Med, 7*(13), 290.

Caballero, B., et al. (2003). *Encyclopedia of food sciences and nutrition* (2nd ed.): Academic Press.

Calabria, A. R., & Shusta, E. V. (2008). A genomic comparison of in vivo and in vitro brain microvascular endothelial cells. *J Cereb Blood Flow Metab, 28*(1), 135-148.

Case, L. E., et al. (2012). Infantile Pompe disease on ERT: update on clinical presentation, musculoskeletal management, and exercise considerations. *Am J Med Genet C Semin Med Genet, 160C*(1), 69-79.

Cencioni, M. T., et al. (2022). Immune Reconstitution Following Autologous Hematopoietic Stem Cell Transplantation for Multiple Sclerosis: A Review on Behalf of the EBMT Autoimmune Diseases Working Party. *Frontiers in Immunology, 12*.

Chakrapani, A., et al. (2010). Treatment of infantile Pompe disease with alglucosidase alpha: the UK experience. *J Inherit Metab Dis, 33*(6), 747-750.

Chamarthy, M. R., et al. (2018). Pulmonary vascular pathophysiology. *Cardiovascular Diagnosis and Therapy, 8*(3), 208-213.

Chamoles, N. A., et al. (2004). Glycogen storage disease type II: enzymatic screening in dried blood spots on filter paper. *Clinica Chimica Acta, 347*(1-2), 97-102.

Chapillon, P., et al. (1998). Early development of synchronized walking on the rotorod in rats - Effects of training and handling. *Behavioural Brain Research, 93*(1-2), 77-81.

Charpentier, E., & Marraffini, L. A. (2014). Harnessing CRISPR-Cas9 immunity for genetic engineering. *Current Opinion in Microbiology, 19*, 114-119.

Chen, J. W., et al. (2017). General principles of binding between cell surface receptors and multi-specific ligands: A computational study. *Plos Computational Biology, 13*(10).

Chen, X., et al. (2013). Fusion protein linkers: property, design and functionality. *Adv Drug Deliv Rev, 65*(10), 1357-1369.

Cheng, C. S., et al. (2014). Conditions that promote primary human skeletal myoblast culture and muscle differentiation in vitro. *American Journal of Physiology-Cell Physiology, 306*(4), C385-C395.

Chichili, V. P. R., et al. (2013). Linkers in the structural biology of protein-protein interactions. *Protein Science, 22*(2), 153-167.

Chikwana, V. M., et al. (2013). Structural basis for 2 '-phosphate incorporation into glycogen by glycogen synthase. *Proceedings of the National Academy of Sciences of the United States of America, 110*(52), 20976-20981.

Chiorini, J. A., et al. (1997). Cloning of adeno-associated virus type 4 (AAV4) and generation of recombinant AAV4 particles. *J Virol, 71*(9), 6823-6833.

Chiu, N. H. L., & Watson, A. L. (2017). Measuring beta-Galactosidase Activity in Gram-Positive Bacteria Using a Whole-Cell Assay with MUG as a Fluorescent Reporter. *Curr Protoc Toxicol, 74*, 4 44 41-44 44 48.

Christensen, C. L., & Choy, F. Y. M. (2017). A Prospective Treatment Option for Lysosomal Storage Diseases: CRISPR/Cas9 Gene Editing Technology for Mutation Correction in Induced Pluripotent Stem Cells. *Diseases, 5*(1).

Clarke, L. A., et al. (2009). Long-term efficacy and safety of laronidase in the treatment of mucopolysaccharidosis I. *Pediatrics, 123*(1), 229-240.

Clausen, M. V., et al. (2017). The Structure and Function of the Na,K-ATPase Isoforms in Health and Disease. *Front Physiol, 8*, 371.

Colella, P., et al. (2019). AAV Gene Transfer with Tandem Promoter Design Prevents Anti-transgene Immunity and Provides Persistent Efficacy in Neonate Pompe Mice. *Mol Ther Methods Clin Dev, 12*, 85-101.

Consiglio, A., et al. (2001). In vivo gene therapy of metachromatic leukodystrophy by lentiviral vectors: correction of neuropathology and protection against learning impairments in affected mice. *Nat Med, 7*(3), 310-316.

Coutinho, M. F., et al. (2016). Less Is More: Substrate Reduction Therapy for Lysosomal Storage Disorders. *Int J Mol Sci, 17*(7).

Coutinho, M. F., et al. (2017). Less Is More: Substrate Reduction Therapy for Lysosomal Storage Disorders (vol 17, 1065, 2016). *International Journal of Molecular Sciences, 18*(1).

Cox, T., et al. (2000). Novel oral treatment of Gaucher's disease with N-butyldeoxynojirimycin (OGT 918) to decrease substrate biosynthesis. *Lancet, 355*(9214), 1481-1485.

Cronin, J., et al. (2005). Altering the tropism of lentiviral vectors through pseudotyping. *Curr Gene Ther, 5*(4), 387-398.

Cupler, E. J., et al. (2012). Consensus treatment recommendations for late-onset Pompe disease. *Muscle Nerve, 45*(3), 319-333.

Dahms, N. M., et al. (2008). Strategies for carbohydrate recognition by the mannose 6-phosphate receptors. *Glycobiology, 18*(9), 664-678.

Dahms, N. M., et al. (1994). The bovine mannose 6-phosphate/insulin-like growth factor II receptor. Localization of the insulin-like growth factor II binding site to domains 5-11. *J Biol Chem, 269*(5), 3802-3809.

Daniel-Moreno, A., et al. (2019). CRISPR/Cas9-modified hematopoietic stem cells-present and future perspectives for stem cell transplantation. *Bone Marrow Transplant, 54*(12), 1940-1950.

Dardis, A., & Buratti, E. (2018). Impact, Characterization, and Rescue of Pre-mRNA Splicing Mutations in Lysosomal Storage Disorders. *Genes (Basel), 9*(2).

Dasouki, M., et al. (2014). Pompe disease: literature review and case series. *Neurol Clin, 32*(3), 751-776, ix.

DeNies, M. S., et al. (2019). Clathrin Heavy Chain Knockdown Impacts CXCR4 Signaling and Post-translational Modification. *Frontiers in Cell and Developmental Biology, 7*.

DeRuisseau, L. R., et al. (2009). Neural deficits contribute to respiratory insufficiency in Pompe disease. *Proceedings of the National Academy of Sciences of the United States of America, 106*(23), 9419-9424.

Desai, A. K., et al. (2019). Characterization of immune response in Cross-Reactive Immunological Material (CRIM)-positive infantile Pompe disease patients treated with enzyme replacement therapy. *Mol Genet Metab Rep, 20*, 100475.

Devi, G. R., et al. (1998). An insulin-like growth factor II (IGF-II) affinity-enhancing domain localized within extracytoplasmic repeat 13 of the IGF-II/mannose 6-phosphate receptor. *Mol Endocrinol, 12*(11), 1661-1672.

Di Domenico, C., et al. (2009). Intracranial gene delivery of LV-NAGLU vector corrects neuropathology in murine MPS IIIB. *Am J Med Genet A, 149A*(6), 1209-1218.

DiMauro, S., & Spiegel, R. (2011). Progress and problems in muscle glycogenoses. *Acta Myol, 30*(2), 96-102.

Disease, A. W. G. o. M. o. P., et al. (2006). Pompe disease diagnosis and management guideline. *Genet Med, 8*(5), 267-288.

Do, H. V., et al. (2019). Challenges in treating Pompe disease: an industry perspective. *Ann Transl Med, 7*(13), 291.

Dockendorff, C., et al. (2015). Indolinyl-Thiazole Based Inhibitors of Scavenger Receptor-BI (SR-BI)-Mediated Lipid Transport. *Acs Medicinal Chemistry Letters, 6*(4), 375-380.

Doerfler, P. A., et al. (2016). Copackaged AAV9 Vectors Promote Simultaneous Immune Tolerance and Phenotypic Correction of Pompe Disease. *Hum Gene Ther, 27*(1), 43-59.

Doudna, J. A., & Charpentier, E. (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science, 346*(6213), 1258096.

Douillard-Guilloux, G., et al. (2008). Modulation of glycogen synthesis by RNA interference: towards a new therapeutic approach for glycogenosis type II. *Hum Mol Genet, 17*(24), 3876-3886.

Doyon, J. B., et al. (2011). Rapid and efficient clathrin-mediated endocytosis revealed in genome-edited mammalian cells. *Nature Cell Biology, 13*(3), 331-U327.

Duque, S., et al. (2009). Intravenous administration of self-complementary AAV9 enables transgene delivery to adult motor neurons. *Mol Ther, 17*(7), 1187-1196.

Ebbink, B. J., et al. (2018). Classic infantile Pompe patients approaching adulthood: a cohort study on consequences for the brain. *Dev Med Child Neurol, 60*(6), 579-586.

Ebbink, B. J., et al. (2016). Cognitive decline in classic infantile Pompe disease: An underacknowledged challenge. *Neurology, 86*(13), 1260-1261.

Eckhardt, E. R. M., et al. (2006). High density lipoprotein endocytosis by scavenger receptor SR-BII is clathrin-dependent and requires a carboxyl-terminal dileucine motif. *Journal of Biological Chemistry, 281*(7), 4348-4353.

Eddy, S. R. (1996). Hidden Markov models. *Curr Opin Struct Biol, 6*(3), 361-365.

El-Gharbawy, A. H., et al. (2011). Expanding the clinical spectrum of late-onset Pompe disease: dilated arteriopathy involving the thoracic aorta, a novel vascular phenotype uncovered. *Mol Genet Metab, 103*(4), 362-366.

Ellingwood, S. S., & Cheng, A. (2018). Biochemical and clinical aspects of glycogen storage diseases. *Journal of Endocrinology, 238*(3), R131-R141.

Escors, D., & Breckpot, K. (2010). Lentiviral vectors in gene therapy: their current status and future potential. *Arch Immunol Ther Exp (Warsz), 58*(2), 107-119.

Falk, D. J., et al. (2015). Comparative impact of AAV and enzyme replacement therapy on respiratory and cardiac function in adult Pompe mice. *Mol Ther Methods Clin Dev, 2*, 15007.

Fassler, R. (2004). Lentiviral transgene vectors - Green light for efficient production of transgenic farm animals. *Embo Reports, 5*(1), 28-29.

Favret, J. M., et al. (2020). Pre-clinical Mouse Models of Neurodegenerative Lysosomal Storage Diseases. *Frontiers in Molecular Biosciences, 7*.

Ferkany, J. W. (1987). The Radioreceptor Assay - a Simple, Sensitive and Rapid Analytical Procedure. *Life Sciences, 41*(7), 881-884.

Ferri, L., et al. (2016). Double-target Antisense U1snRNAs Correct Mis-splicing Due to c.639+861C>T and c.639+919G>A GLA Deep Intronic Mutations. *Mol Ther Nucleic Acids, 5*(10), e380.

Finkel, R. S., et al. (2016). Treatment of infantile-onset spinal muscular atrophy with nusinersen: a phase 2, open-label, dose-escalation study. *Lancet, 388*(10063), 3017-3026.

Fraites, T. J., Jr., et al. (2002). Correction of the enzymatic and functional deficits in a model of Pompe disease using adeno-associated virus vectors. *Mol Ther, 5*(5 Pt 1), 571-578.

Fraldi, A., et al. (2018). Gene therapy for mucopolysaccharidoses: in vivo and ex vivo approaches. *Ital J Pediatr, 44*(Suppl 2), 130.

Fratantoni, J. C., et al. (1968). Hurler and Hunter syndromes: mutual correction of the defect in cultured fibroblasts. *Science, 162*(3853), 570-572.

Fukuda, T., et al. (2006). Dysfunction of endocytic and autophagic pathways in a lysosomal storage disease. *Ann Neurol, 59*(4), 700-708.

Fuller, M., et al. (2006). Epidemiology of lysosomal storage diseases: an overview. In A. Mehta, M. Beck, & G. Sunder-Plassmann (Eds.), *Fabry Disease: Perspectives from 5 Years of FOS*. Oxford.

Galej, W. P., et al. (2014). Structural studies of the spliceosome: zooming into the heart of the machine. *Curr Opin Struct Biol, 25*, 57-66.

Gao, G., et al. (2005). New recombinant serotypes of AAV vectors. *Curr Gene Ther, 5*(3), 285-297.

Gao, X., et al. (2014). Effective dystrophin restoration by a novel muscle-homing peptide-morpholino conjugate in dystrophin-deficient mdx mice. *Mol Ther, 22*(7), 1333-1341.

Garcia-Contreras, R., et al. (2012). Why in vivo may not equal in vitro - new effectors revealed by measurement of enzymatic activities under the same in vivo-like assay conditions. *FEBS J, 279*(22), 4145-4159.

Garcia-Perez, L., et al. (2020). Successful Preclinical Development of Gene Therapy for Recombinase-Activating Gene-1-Deficient SCID. *Molecular Therapy-Methods & Clinical Development, 17*, 666-682.

Gary-Bobo, M., et al. (2007). Mannose 6-phosphate receptor targeting and its applications in human diseases. *Current Medicinal Chemistry, 14*(28), 2945-2953.

Gelens, L., & Saurin, A. T. (2018). Exploring the Function of Dynamic Phosphorylation-Dephosphorylation Cycles. *Developmental Cell, 44*(6), 659-663.

Geljic, I. S., et al. (2021). Viral Interactions with Adaptor-Protein Complexes: A Ubiquitous Trait among Viral Species. *International Journal of Molecular Sciences, 22*(10).

Germain, D. P., et al. (2016). Treatment of Fabry's Disease with the Pharmacologic Chaperone Migalastat. *N Engl J Med, 375*(6), 545-555.

Ghosh, P., et al. (2003). Mannose 6-phosphate receptors: new twists in the tale. *Nat Rev Mol Cell Biol, 4*(3), 202-212.

Ghosh, S., et al. (2020). Viral Vector Systems for Gene Therapy: A Comprehensive Literature Review of Progress and Biosafety Challenges. *Applied Biosafety, 25*(1), 7-18.

Gibbons, B. J., et al. (2002). Crystal structure of the autocatalytic initiator of glycogen biosynthesis, glycogenin. *Journal of Molecular Biology, 319*(2), 463-477.

Girod, A., et al. (2002). The VP1 capsid protein of adeno-associated virus type 2 is carrying a phospholipase A2 domain required for virus infectivity. *J Gen Virol, 83*(Pt 5), 973-978.

Goemans, C., et al. (2014). Folding mechanisms of periplasmic proteins. *Biochim Biophys Acta, 1843*(8), 1517-1528.

Goina, E., et al. (2017). Glycogen Reduction in Myotubes of Late-Onset Pompe Disease Patients Using Antisense Technology. *Mol Ther, 25*(9), 2117-2128.

Goldberg, L. R., et al. (2014). The murine long-term multi-lineage renewal marrow stem cell is a cycling cell. *Leukemia, 28*(4), 813-822.

Goldstein, J. L., & Brown, M. S. (2009). The LDL Receptor. *Arteriosclerosis Thrombosis and Vascular Biology, 29*(4), 431-438.

Grabowski, G. A. (2012). Gaucher disease and other storage disorders. *Hematology Am Soc Hematol Educ Program, 2012*, 13-18.

Gray, S. J., et al. (2010). Viral vectors and delivery strategies for CNS gene therapy. *Ther Deliv, 1*(4), 517-534.

Gritti, A. (2011). Gene therapy for lysosomal storage disorders. *Expert Opin Biol Ther, 11*(9), 1153-1167.

Guler-Gane, G., et al. (2016). Overcoming the Refractory Expression of Secreted Recombinant Proteins in Mammalian Cells through Modification of the Signal Peptide and Adjacent Amino Acids. *PLoS One, 11*(5), e0155340.

Gungor, D., et al. (2011). Survival and associated factors in 268 adults with Pompe disease prior to treatment with enzyme replacement therapy. *Orphanet Journal of Rare Diseases, 6*.

Hafemeister, C., & Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biology, 20*(1).

Hagemans, M. L., et al. (2005). Clinical manifestation and natural course of late-onset Pompe's disease in 54 Dutch patients. *Brain, 128*(Pt 3), 671-677.

Haltia, M. (2006). The neuronal ceroid-lipofuscinoses: From past to present. *Biochimica Et Biophysica Acta-Molecular Basis of Disease, 1762*(10), 850-856.

Haneef, S. A. S., & Doss, C. G. P. (2016). Personalized Pharmacoperones for Lysosomal Storage Disorder: Approach for Next-Generation Treatment. *Advances in Protein Chemistry and Structural Biology, Vol 102: Personalized Medicine, 102*, 225-265.

Hannoun, Z., et al. (2010). Post-translational modification by SUMO. *Toxicology, 278*(3), 288-293.

Hansen, C. G., & Nichols, B. J. (2009). Molecular mechanisms of clathrin-independent endocytosis. *Journal of Cell Science, 122*(11), 1713-1721.

Hasilik, A., & Neufeld, E. F. (1980). Biosynthesis of lysosomal enzymes in fibroblasts. Phosphorylation of mannose residues. *J Biol Chem, 255*(10), 4946-4950.

Hatse, S., et al. (2002). Chemokine receptor inhibition by AMD3100 is strictly confined to CXCR4. *Febs Letters, 527*(1-3), 255-262.

Hermans, M. M., et al. (2004). Twenty-two novel mutations in the lysosomal alpha-glucosidase gene (GAA) underscore the genotype-phenotype correlation in glycogen storage disease type II. *Hum Mutat, 23*(1), 47-56.

Hermans, M. M., et al. (1993). Human lysosomal alpha-glucosidase: functional characterization of the glycosylation sites. *Biochem J, 289 ( Pt 3)*, 681-686.

Hers, H. G. (1963). Alpha-Glucosidase Deficiency in Generalized Glycogen-Storage Disease (Pompes Disease). *Biochemical Journal, 86*(1), 11-&.

Herzog, A., et al. (2012). A cross-sectional single-centre study on the spectrum of Pompe disease, German patients: molecular analysis of the GAA gene, manifestation and genotype-phenotype correlations. *Orphanet J Rare Dis, 7*, 35.

Hille, F., & Charpentier, E. (2016). CRISPR-Cas: biology, mechanisms and relevance. *Philosophical Transactions of the Royal Society B-Biological Sciences, 371*(1707).

Hille, F., et al. (2018). The Biology of CRISPR-Cas: Backward and Forward. *Cell, 172*(6), 1239-1259. doi:10.1016/j.cell.2017.11.032

Hobbs, J. R., et al. (1981). Reversal of clinical features of Hurler's disease and biochemical improvement after treatment by bone-marrow transplantation. *Lancet, 2*(8249), 709-712.

Hocquemiller, M., et al. (2016). Adeno-Associated Virus-Based Gene Therapy for CNS Diseases. *Hum Gene Ther, 27*(7), 478-496.

Hollak, C. E., et al. (2009). Miglustat (Zavesca) in type 1 Gaucher disease: 5-year results of a post-authorisation safety surveillance programme. *Pharmacoepidemiol Drug Saf, 18*(9), 770-777.

Huang, D. B., et al. (2021). Synthetic biology approaches in regulation of targeted gene expression. *Current Opinion in Plant Biology, 63*.

Huang, J. Y., et al. (2020). CRISPR-Cas9 generated Pompe knock-in murine model exhibits early-onset hypertrophic cardiomyopathy and skeletal muscle weakness. *Sci Rep, 10*(1), 10321.

Huang, Z., et al. (2016). A study on the effects of linker flexibility on acid phosphatase PhoC-GFP fusion protein using a novel linker library. *Enzyme Microb Technol, 83*, 1-6.

Hwu, W. L., & Chien, Y. H. (2020). Development of Newborn Screening for Pompe Disease. *Int J Neonatal Screen, 6*(1), 5.

Jaganathan, K., et al. (2019). Predicting Splicing from Primary Sequence with Deep Learning. *Cell, 176*(3), 535-548 e524.

Janiszewski, L. N., et al. (2021). Characterization of Global Gene Expression, Regulation of Metal Ions, and Infection Outcomes in Immune-Competent 129S6 Mouse Macrophages. *Infection and Immunity, 89*(11).

Jirka, S. M. G., et al. (2018). Cyclic Peptides to Improve Delivery and Exon Skipping of Antisense Oligonucleotides in a Mouse Model for Duchenne Muscular Dystrophy. *Molecular Therapy, 26*(1), 132-147.

Johnson, L. N. (1992). Glycogen-Phosphorylase - Control by Phosphorylation and Allosteric Effectors. *Faseb Journal, 6*(6), 2274-2282.

Kadlecova, Z., et al. (2017). Regulation of clathrin-mediated endocytosis by hierarchical allosteric activation of AP2. *Journal of Cell Biology, 216*(1), 167-179.

Kaksonen, M., & Roux, A. (2018). Mechanisms of clathrin-mediated endocytosis. *Nature Reviews Molecular Cell Biology, 19*(5), 313-326.

Kan, S. H., et al. (2014). Insulin-like growth factor II peptide fusion enables uptake and lysosomal delivery of alpha-N-acetylglucosaminidase to mucopolysaccharidosis type IIIB fibroblasts. *Biochem J, 458*(2), 281-289.

Kang, J. X., et al. (1997). Mannose-6-phosphate/insulin-like growth factor-II receptor is a receptor for retinoic acid. *Proc Natl Acad Sci U S A, 94*(25), 13671-13676.

Kang, Y., et al. (2002). In vivo gene transfer using a nonprimate lentiviral vector pseudotyped with Ross River Virus glycoproteins. *J Virol, 76*(18), 9378-9388.

Kattenhorn, L. M., et al. (2016). Adeno-Associated Virus Gene Therapy for Liver Disease. *Hum Gene Ther, 27*(12), 947-961.

Katz, R. A., & Skalka, A. M. (1994). The retroviral enzymes. *Annu Rev Biochem, 63*, 133-173.

Kay, M. A., et al. (2001). Viral vectors for gene therapy: the art of turning infectious agents into vehicles of therapeutics. *Nat Med, 7*(1), 33-40.

Keeling, K. M., & Bedwell, D. M. (2011). Suppression of nonsense mutations as a therapeutic approach to treat genetic diseases. *Wiley Interdiscip Rev RNA, 2*(6), 837-852.

Khanna, A., et al. (2000). Effects of liver-derived dendritic cell progenitors on Th1-and Th2-like cytokine responses in vitro and in vivo. *Journal of Immunology, 164*(3), 1346-1354.

Kilic, T., et al. (2000). Intracranial inhibition of platelet-derived growth factor-mediated glioblastoma cell growth by an orally active kinase inhibitor of the 2-phenylaminopyrimidine class. *Cancer Research, 60*(18), 5143-5150.

Kishnani, P. S., & Beckemeyer, A. A. (2014). New therapeutic approaches for Pompe disease: enzyme replacement therapy and beyond. *Pediatr Endocrinol Rev, 12 Suppl 1*, 114-124.

Kishnani, P. S., et al. (2010). Cross-reactive immunologic material status affects treatment outcomes in Pompe disease infants. *Mol Genet Metab, 99*(1), 26-33.

Kishnani, P. S., et al. (2006). A retrospective, multinational, multicenter study on the natural history of infantile-onset Pompe disease. *J Pediatr, 148*(5), 671-676.

Kishnani, P. S., & Koeberl, D. D. (2019). Liver depot gene therapy for Pompe disease. *Ann Transl Med, 7*(13), 288.

Kishnani, P. S., et al. (2006). Pompe disease diagnosis and management guideline. *Genetics in Medicine, 8*(5), 267-288.

Klein, J. S., et al. (2014). Design and characterization of structured protein linkers with differing flexibilities. *Protein Eng Des Sel, 27*(10), 325-330.

Kobayashi, H., et al. (2005). Neonatal gene therapy of MPS I mice by intravenous injection of a lentiviral vector. *Mol Ther, 11*(5), 776-789.

Kobayashi, H., et al. (2010). Prognostic factors for the late onset Pompe disease with enzyme replacement therapy: from our experience of 4 cases including an autopsy case. *Mol Genet Metab, 100*(1), 14-19.

Koeberl, D. D., et al. (2011a). Enhanced efficacy of enzyme replacement therapy in Pompe disease through mannose-6-phosphate receptor expression in skeletal muscle. *Mol Genet Metab, 103*(2), 107-112.

Koeberl, D. D., et al. (2011b). Enhanced efficacy of enzyme replacement therapy in Pompe disease through mannose-6-phosphate receptor expression in skeletal muscle. *Molecular Genetics and Metabolism, 103*(2), 107-112.

Kollberg, G., et al. (2007). Cardiomyopathy and exercise intolerance in muscle glycogen storage disease 0. *N Engl J Med, 357*(15), 1507-1514.

Korlimarla, A., et al. (2019). An emerging phenotype of central nervous system involvement in Pompe disease: from bench to bedside and beyond. *Annals of Translational Medicine, 7*(13).

Krause, D. S., et al. (2001). Multi-organ, multi-lineage engraftment by a single bone marrow-derived stem cell. *Cell, 105*(3), 369-377.

Kroos, M., et al. (2012). Update of the pompe disease mutation database with 60 novel GAA sequence variants and additional studies on the functional effect of 34 previously reported variants. *Hum Mutat, 33*(8), 1161-1165.

Kroos, M., et al. (2008). Update of the Pompe disease mutation database with 107 sequence variants and a format for severity rating. *Hum Mutat, 29*(6), E13-26.

Kumari, S., et al. (2010). Endocytosis unplugged: multiple ways to enter the cell. *Cell Research, 20*(3), 256-275.

Kyosen, S. O., et al. (2010). Neonatal gene transfer using lentiviral vector for murine Pompe disease: long-term expression and glycogen reduction. *Gene Ther, 17*(4), 521-530.

Lamsfus-Calle, A., et al. (2020). Hematopoietic stem cell gene therapy: The optimal use of lentivirus and gene editing approaches. *Blood Rev, 40*, 100641.

Langan, T. J., et al. (2016). Evidence for improved survival in postsymptomatic stem cell-transplanted patients with Krabbe's disease. *J Neurosci Res, 94*(11), 1189-1194.

LeBowitz, J. H., et al. (2004). Glycosylation-independent targeting enhances enzyme delivery to lysosomes and decreases storage in mucopolysaccharidosis type VII mice. *Proc Natl Acad Sci U S A, 101*(9), 3083-3088.

Lenahan, C., et al. (2019). Scavenger Receptor Class B type 1 (SR-B1) and the modifiable risk factors of stroke. *Chin Neurosurg J, 5*, 30.

Li, M. D. (2018). Enzyme Replacement Therapy: A Review and Its Role in Treating Lysosomal Storage Diseases. *Pediatric Annals, 47*(5), E191-E197.

Li, Y., et al. (2004). Direct multiplex assay of lysosomal enzymes in dried blood spots for newborn screening. *Clin Chem, 50*(10), 1785-1796.

Lim, J. A., et al. (2014a). Pompe disease: from pathophysiology to therapy and back again. *Front Aging Neurosci, 6*, 177.

Lim, J. A., et al. (2014b). Pompe disease: from pathophysiology to therapy and back again. *Frontiers in Aging Neuroscience, 6*.

Lim, J. A., et al. (2019). Intravenous Injection of an AAV-PHP.B Vector Encoding Human Acid alpha-Glucosidase Rescues Both Muscle and CNS Defects in Murine Pompe Disease. *Mol Ther Methods Clin Dev, 12*, 233-245.

Lin, D., et al. (2005). AAV2/5 vector expressing galactocerebrosidase ameliorates CNS disease in the murine model of globoid-cell leukodystrophy more efficiently than AAV2. *Mol Ther, 12*(3), 422-430.

Lin, S. X., et al. (2004). Endocytosed cation-independent mannose 6-phosphate receptor traffics via the endocytic recycling compartment en route to the trans-Golgi network and a subpopulation of late endosomes. *Molecular Biology of the Cell, 15*(2), 721-733.

LoDuca, P. A., et al. (2009). Hepatic gene transfer as a means of tolerance induction to transgene products. *Curr Gene Ther, 9*(2), 104-114.

Lorente, S., et al. (2020). The liver, a functionalized vascular structure. *Scientific Reports, 10*(1).

Lu, I. L., et al. (2003). Correction/mutation of acid alpha-D-glucosidase gene by modified single-stranded oligonucleotides: in vitro and in vivo studies. *Gene Ther, 10*(22), 1910-1916.

Luth, S., et al. (2008). Ectopic expression of neural autoantigen in mouse liver suppresses experimental autoimmune neuroinflammation by inducing antigen-specific Tregs. *Journal of Clinical Investigation, 118*(10), 3403-3410.

Lutzko, C., et al. (2003). Lentivirus vectors incorporating the immunoglobulin heavy chain enhancer and matrix attachment regions provide position-independent expression in B lymphocytes. *J Virol, 77*(13), 7341-7351.

Lyseng-Williamson, K. A. (2014). Miglustat: a review of its use in Niemann-Pick disease type C. *Drugs, 74*(1), 61-74.

Macauley, S. L. (2016). Combination Therapies for Lysosomal Storage Diseases: A Complex Answer to a Simple Problem. *Pediatr Endocrinol Rev, 13 Suppl 1*, 639-648.

Maga, J. A., et al. (2013). Glycosylation-independent lysosomal targeting of acid alpha-glucosidase enhances muscle glycogen clearance in pompe mice. *J Biol Chem, 288*(3), 1428-1438.

Mah, C. S., et al. (2010). Gel-mediated delivery of AAV1 vectors corrects ventilatory function in Pompe mice with established disease. *Mol Ther, 18*(3), 502-510.

Malgieri, A., et al. (2010). Bone marrow and umbilical cord blood human mesenchymal stem cells: state of the art. *Int J Clin Exp Med, 3*(4), 248-269.

Malicdan, M. C., & Nishino, I. (2012). Autophagy in lysosomal myopathies. *Brain Pathol, 22*(1), 82-88.

Manganelli, F., & Ruggiero, L. (2013). Clinical features of Pompe disease. *Acta Myol, 32*(2), 82-84.

Markham, A. (2016). Migalastat: First Global Approval. *Drugs, 76*(11), 1147-1152.

Marques, A. R. A., & Saftig, P. (2019). Lysosomal storage disorders - challenges, concepts and avenues for therapy: beyond rare diseases. *J Cell Sci, 132*(2).

Martarello, R. D., et al. (2019). Optimization and partial purification of beta-galactosidase production by Aspergillus niger isolated from Brazilian soils using soybean residue. *Amb Express, 9*.

Martin-Touaux, E., et al. (2002). Muscle as a putative producer of acid alpha-glucosidase for glycogenosis type II gene therapy. *Hum Mol Genet, 11*(14), 1637-1645.

Martinez-Molina, E., et al. (2020). Large-Scale Production of Lentiviral Vectors: Current Perspectives and Challenges. *Pharmaceutics, 12*(11).

Martiniuk, F., et al. (1986). Isolation of a cDNA for human acid alpha-glucosidase and detection of genetic heterogeneity for mRNA in three alpha-glucosidase-deficient patients. *Proc Natl Acad Sci U S A, 83*(24), 9641-9644.

Massaro, G., et al. (2021). Gene Therapy for Lysosomal Storage Disorders: Ongoing Studies and Clinical Development. *Biomolecules, 11*(4).

McBurney, M. W., et al. (1991). The mouse Pgk-1 gene promoter contains an upstream activator sequence. *Nucleic Acids Res, 19*(20), 5755-5761.

McCafferty, E. H., & Scott, L. J. (2019). Migalastat: A Review in Fabry Disease. *Drugs, 79*(5), 543-554.

McLaughlin, S. K., et al. (1988). Adeno-associated virus general transduction vectors: analysis of proviral structures. *J Virol, 62*(6), 1963-1973.

McPherson, P. S., et al. (2009). Clathrin-Mediated Endocytosis. *Trafficking inside Cells: Pathways, Mechanisms and Regulation*, 159-182.

Medina, D. L., & Ballabio, A. (2015). Lysosomal calcium regulates autophagy. *Autophagy, 11*(6), 970-971.

Meikle, P. J., et al. (1999). Prevalence of lysosomal storage disorders. *JAMA, 281*(3), 249-254.

Mercier, E., et al. (2017). Signal recognition particle binds to translating ribosomes before emergence of a signal anchor sequence. *Nucleic Acids Res, 45*(20), 11858-11866.

Merten, O. W., et al. (2016). Production of lentiviral vectors. *Mol Ther Methods Clin Dev, 3*, 16017.

Mettlen, M., et al. (2018). Regulation of Clathrin-Mediated Endocytosis. *Annu Rev Biochem, 87*, 871-896.

Metzler, M., et al. (2001). HIP1 functions in clathrin-mediated endocytosis through binding to clathrin and adaptor protein 2. *Journal of Biological Chemistry, 276*(42), 39271-39276.

Milone, M. C., & O'Doherty, U. (2018). Clinical use of lentiviral vectors. *Leukemia, 32*(7), 1529-1541.

Mingozzi, F., & High, K. A. (2013). Immune responses to AAV vectors: overcoming barriers to successful gene therapy. *Blood, 122*(1), 23-36.

Mingozzi, F., et al. (2003). Induction of immune tolerance to coagulation factor IX antigen by in vivo hepatic gene transfer. *Journal of Clinical Investigation, 111*(9), 1347-1356.

Mokuno, H., et al. (1994). Effect of the 39-Kda Receptor-Associated Protein on the Hepatic-Uptake and Endocytosis of Chylomicron Remnants and Low-Density Lipoproteins in the Rat. *Journal of Biological Chemistry, 269*(18), 13238-13243.

Montiel-Equihua, C. A., et al. (2012). The beta-globin locus control region in combination with the EF1alpha short promoter allows enhanced lentiviral vector-mediated erythroid gene expression with conserved multilineage activity. *Mol Ther, 20*(7), 1400-1409.

Montini, E., et al. (2006). Hematopoietic stem cell gene transfer in a tumor-prone mouse model uncovers low genotoxicity of lentiviral vector integration. *Nat Biotechnol, 24*(6), 687-696.

Moreland, R. J., et al. (2005). Lysosomal acid alpha-glucosidase consists of four different peptides processed from a single chain precursor. *J Biol Chem, 280*(8), 6780-6791.

Mueller, P., et al. (2013). Cardiac disease in children and young adults with various lysosomal storage diseases: Comparison of echocardiographic and ECG changes among clinical groups. *Int J Cardiol Heart Vessel, 2*, 1-7.

Muses, S., et al. (2011). A New Extensively Characterised Conditionally Immortal Muscle Cell-Line for Investigating Therapeutic Strategies in Muscular Dystrophies. *PLoS One, 6*(9).

Nakagawa, T. Y., & Rudensky, A. Y. (1999). The role of lysosomal proteinases in MHC class II-mediated antigen processing and presentation. *Immunol Rev, 172*, 121-129.

Nakai, H., et al. (2001). Extrachromosomal recombinant adeno-associated virus vector genomes are primarily responsible for stable liver transduction in vivo. *J Virol, 75*(15), 6969-6976.

Neufeld, E. F. (1991). Lysosomal storage diseases. *Annu Rev Biochem, 60*, 257-280.

Ng, A. P., & Alexander, W. S. (2017). Haematopoietic stem cells: past, present and future. *Cell Death Discov, 3*, 17002.

Nicolino, M., et al. (1997). Glycogen-storage disease type II (acid maltase deficiency): identification of a novel small deletion (delCC482+483) in French patients. *Biochem Biophys Res Commun, 235*(1), 138-141.

Nieuwenhuis, B., et al. (2021). Optimization of adeno-associated viral vector-mediated transduction of the corticospinal tract: comparison of four promoters. *Gene Ther, 28*(1-2), 56-74.

Nieuwenhuis, B., et al. (2023). Improving adeno-associated viral (AAV) vector-mediated transgene expression in retinal ganglion cells: comparison of five promoters. *Gene Ther*.

Nino, M. Y., et al. (2019). Extension of the Pompe mutation database by linking disease-associated variants to clinical severity. *Hum Mutat, 40*(11), 1954-1967.

Nonnenmacher, M., & Weber, T. (2012). Intracellular transport of recombinant adeno-associated virus vectors. *Gene Ther, 19*(6), 649-658.

Nykjaer, A., et al. (1998). Mannose 6-phosphate/insulin-like growth factor-II receptor targets the urokinase receptor to lysosomes via a novel binding interaction. *J Cell Biol, 141*(3), 815-828.

Orchard, P. J., & Tolar, J. (2010). Transplant outcomes in leukodystrophies. *Semin Hematol, 47*(1), 70-78.

Orth, M., & Mundegar, R. R. (2003). Effect of acid maltase deficiency on the endosomal/lysosomal system and glucose transporter 4. *Neuromuscul Disord, 13*(1), 49-54.

Papasaikas, P., & Valcarcel, J. (2016). The Spliceosome: The Ultimate RNA Chaperone and Sculptor. *Trends Biochem Sci, 41*(1), 33-45.

Papayannakos, C., & Daniel, R. (2013). Understanding lentiviral vector chromatin targeting: working to reduce insertional mutagenic potential for gene therapy. *Gene Therapy, 20*(6), 581-588.

Parenti, G., et al. (2015). Lysosomal storage diseases: from pathophysiology to therapy. *Annu Rev Med, 66*, 471-486.

Parenti, G., et al. (2015). Pharmacological Chaperone Therapy: Preclinical Development, Clinical Translation, and Prospects for the Treatment of Lysosomal Storage Disorders. *Mol Ther, 23*(7), 1138-1148.

Pascual, J. M., & Roe, C. R. (2013). Systemic metabolic abnormalities in adult-onset acid maltase deficiency: beyond muscle glycogen accumulation. *JAMA Neurol, 70*(6), 756-763.

Passini, M. A., et al. (2005). AAV vector-mediated correction of brain pathology in a mouse model of Niemann-Pick A disease. *Mol Ther, 11*(5), 754-762.

Passini, M. A., et al. (2003). Intraventricular brain injection of adeno-associated virus type 1 (AAV1) in neonatal mice results in complementary patterns of neuronal transduction to AAV2 and total long-term correction of storage lesions in the brains of beta-glucuronidase-deficient mice. *J Virol, 77*(12), 7034-7040.

Patterson, M. C., et al. (2007). Miglustat for treatment of Niemann-Pick C disease: a randomised controlled study. *Lancet Neurol, 6*(9), 765-772.

Pauwels, K., et al. (2009). State-of-the-art lentiviral vectors for research use: risk assessment and biosafety recommendations. *Curr Gene Ther, 9*(6), 459-474.

Pena, L. D. M., et al. (2019). Safety, tolerability, pharmacokinetics, pharmacodynamics, and exploratory efficacy of the novel enzyme replacement therapy avalglucosidase alfa (neoGAA) in treatment-naive and alglucosidase alfa-treated patients with late-onset Pompe disease: A phase 1, open-label, multicenter, multinational, ascending dose study. *Neuromuscul Disord, 29*(3), 167-186.

Pennazio, S. (2007). Genetics and virology: two interdisciplinary branches of biology. *Riv Biol, 100*(1), 119-146.

Pera, E. M., et al. (2001). Neural and head induction by insulin-like growth factor signals. *Dev Cell, 1*(5), 655-665.

Peruzzo, P., et al. (2019). Molecular genetics of Pompe disease: a comprehensive overview. *Ann Transl Med, 7*(13), 278.

Philp, A., et al. (2012). More than a store: regulatory roles for glycogen in skeletal muscle adaptation to exercise. *American Journal of Physiology-Endocrinology and Metabolism, 302*(11), E1343-E1351.

Pineda, M., et al. (2018). Miglustat in Niemann-Pick disease type C patients: a review. *Orphanet J Rare Dis, 13*(1), 140.

Piotrowska, E., et al. (2006). Genistein-mediated inhibition of glycosaminoglycan synthesis as a basis for gene expression-targeted isoflavone therapy for mucopolysaccharidoses. *Eur J Hum Genet, 14*(7), 846-852.

Piras, G., et al. (2020). Lentiviral Hematopoietic Stem Cell Gene Therapy Rescues Clinical Phenotypes in a Murine Model of Pompe Disease. *Molecular Therapy-Methods & Clinical Development, 18*, 558-570.

Piras, G. M.-E., C. Chan, A. Wantuch, S. (2020). Lentiviral vector mediated haematopoietic stem cell gene therapy overexpressing acid-alpha-glucosidase rescues the clinical phenotypes in a murine model of Pompe disease. *Human Gene Therapy,*.

Pittis, M. G., et al. (2008). Molecular and functional characterization of eight novel GAA mutations in Italian infants with Pompe disease. *Hum Mutat, 29*(6), E27-36.

Platt, F. M., et al. (2018). Lysosomal storage diseases. *Nat Rev Dis Primers, 4*(1), 27.

Platt, F. M., & Lachmann, R. H. (2009). Treating lysosomal storage disorders: current practice and future prospects. *Biochim Biophys Acta, 1793*(4), 737-745.

Pluta, K., & Kacprzak, M. M. (2009). Use of HIV as a gene transfer vector. *Acta Biochim Pol, 56*(4), 531-595.

Poe, M. D., et al. (2014). Early treatment is associated with improved cognition in Hurler syndrome. *Ann Neurol, 76*(5), 747-753.

Poletto, E., et al. (2020). Genome Editing for Mucopolysaccharidoses. *Int J Mol Sci, 21*(2).

Poole, R. M. (2014). Eliglustat: first global approval. *Drugs, 74*(15), 1829-1836.

Porteus, M. (2011). Translating the lessons from gene therapy to the development of regenerative medicine. *Mol Ther, 19*(3), 439-441.

Porto, C., et al. (2009). The pharmacological chaperone N-butyldeoxynojirimycin enhances enzyme replacement therapy in Pompe disease fibroblasts. *Mol Ther, 17*(6), 964-971.

Preisler, N., et al. (2013). Late-onset Pompe disease is prevalent in unclassified limb-girdle muscular dystrophies. *Mol Genet Metab, 110*(3), 287-289.

Preiss, J. (2019). Glycogen Biosynthesis☆. In T. M. Schmidt (Ed.), *Encyclopedia of Microbiology (Fourth Edition)* (pp. 443-456). Oxford: Academic Press.

Priller, J., et al. (2001). Targeting gene-modified hematopoietic cells to the central nervous system: use of green fluorescent protein uncovers microglial engraftment. *Nat Med, 7*(12), 1356-1361.

Puertollano, R., & Raben, N. (2018). Pompe disease: how to solve many problems with one solution. *Ann Transl Med, 6*(15), 313.

Puzzo, F., et al. (2017). Rescue of Pompe disease in mice by AAV-mediated liver delivery of secretable acid alpha-glucosidase. *Sci Transl Med, 9*(418).

Qiu, K., et al. (2012). Spinal delivery of AAV vector restores enzyme activity and increases ventilation in Pompe mice. *Mol Ther, 20*(1), 21-27.

Raben, N., et al. (2008). Suppression of autophagy in skeletal muscle uncovers the accumulation of ubiquitinated proteins and their potential role in muscle damage in Pompe disease. *Hum Mol Genet, 17*(24), 3897-3908.

Raben, N., et al. (1998). Targeted disruption of the acid alpha-glucosidase gene in mice causes an illness with critical features of both infantile and adult human glycogen storage disease type II. *J Biol Chem, 273*(30), 19086-19092.

Raben, N., et al. (2012). Autophagy and mitochondria in Pompe disease: nothing is so new as what has long been forgotten. *Am J Med Genet C Semin Med Genet, 160C*(1), 13-21.

Rakhit, S., et al. (2000). The platelet-derived growth factor receptor stimulation of p42/p44 mitogen-activated protein kinase in airway smooth muscle involves a G-protein-mediated tyrosine phosphorylation of Gab1. *Molecular Pharmacology, 58*(2), 413-420.

Ranieri, M., et al. (2016). Fabry Disease: Recognition, Diagnosis, and Treatment of Neurological Features. *Curr Treat Options Neurol, 18*(7), 33. doi:10.1007/s11940-016-0414-5

Rappoport, J. Z., & Simon, S. M. (2003). Real-time analysis of clathrin-mediated endocytosis during cell migration. (vol 116, pg 847, 2003). *Journal of Cell Science, 116*(6), 1151-1151.

Ratajczak, M. Z. (2008). Phenotypic and functional characterization of hematopoietic stem cells. *Curr Opin Hematol, 15*(4), 293-300.

Rein, D. T., et al. (2003). Evaluation of different tissue specific promoters for cervical cancer gene therapy. *Molecular Therapy, 7*(5), S297-S297.

Rennick, J. J., et al. (2021). Key principles and methods for studying the endocytosis of biological and nanoparticle therapeutics. *Nature Nanotechnology, 16*(3), 266-276.

Reuser, A. J. J., et al. (2019). GAA variants and phenotypes among 1,079 patients with Pompe disease: Data from the Pompe Registry. *Hum Mutat, 40*(11), 2146-2164.

Rey, F. A. (2007). Virology: holed up in a natural crystal. *Nature, 446*(7131), 35-37.

Ritter, B., et al. (2011). Regulating AP-2 interactions to control vesicle size and number during clathrin-mediated endocytosis. *Molecular Biology of the Cell, 22*.

Rohrbach, M., & Clarke, J. T. (2007). Treatment of lysosomal storage disorders : progress with enzyme replacement therapy. *Drugs, 67*(18), 2697-2716.

Roig-Zamboni, V., et al. (2017). Structure of human lysosomal acid alpha-glucosidase-a guide for the treatment of Pompe disease. *Nat Commun, 8*(1), 1111.

Ronzitti, G., et al. (2019). Progress and challenges of gene therapy for Pompe disease. *Ann Transl Med, 7*(13), 287.

Rosenbaum, D. M., et al. (2009). The structure and function of G-protein-coupled receptors. *Nature, 459*(7245), 356-363.

Ruscetti, S. K., et al. (1990). Friend spleen focus-forming virus induces factor independence in an erythropoietin-dependent erythroleukemia cell line. *J Virol, 64*(3), 1057-1062.

Rutherford, T. R., et al. (1979). K562 human leukaemic cells synthesise embryonic haemoglobin in response to haemin. *Nature, 280*(5718), 164-165.

Saibil, H. (2013). Chaperone machines for protein folding, unfolding and disaggregation. *Nat Rev Mol Cell Biol, 14*(10), 630-642.

Saich, R., et al. (2020). Is Newborn Screening the Ultimate Strategy to Reduce Diagnostic Delays in Pompe Disease? The Parent and Patient Perspective. *Int J Neonatal Screen, 6*(1), 1.

Sakaguchi, S., et al. (2008). Regulatory T cells and immune tolerance. *Cell, 133*(5), 775-787.

Salabarria, S. M., et al. (2020). Advancements in AAV-mediated Gene Therapy for Pompe Disease. *J Neuromuscul Dis, 7*(1), 15-31.

Samoylova, T. I., & Smith, B. F. (1999). Elucidation of muscle-binding peptides by phage display screening. *Muscle Nerve, 22*(4), 460-466.

Samulski, R. J., et al. (1989). Helper-free stocks of recombinant adeno-associated viruses: normal integration does not require viral gene expression. *J Virol, 63*(9), 3822-3828.

Samulski, R. J., & Muzyczka, N. (2014). AAV-Mediated Gene Therapy for Research and Therapeutic Purposes. *Annu Rev Virol, 1*(1), 427-451.

Sands, M. S., & Davidson, B. L. (2006). Gene therapy for lysosomal storage diseases. *Mol Ther, 13*(5), 839-849.

Sawado, T., et al. (2003). The beta -globin locus control region (LCR) functions primarily by enhancing the transition from transcription initiation to elongation. *Genes Dev, 17*(8), 1009-1018.

Sawamoto, K., et al. (2018). Gene therapy for Mucopolysaccharidoses. *Mol Genet Metab, 123*(2), 59-68.

Sawen, P., et al. (2018). Murine HSCs contribute actively to native hematopoiesis but with reduced differentiation capacity upon aging. *Elife, 7*.

Schambach, A., et al. (2006). Equal potency of gammaretroviral and lentiviral SIN vectors for expression of O6-methylguanine-DNA methyltransferase in hematopoietic cells. *Mol Ther, 13*(2), 391-400.

Schiffmann, R., et al. (2018). Migalastat improves diarrhea in patients with Fabry disease: clinical-biomarker correlations from the phase 3 FACETS trial. *Orphanet J Rare Dis, 13*(1), 68.

Schoser, B., et al. (2017). Survival and long-term outcomes in late-onset Pompe disease following alglucosidase alfa treatment: a systematic review and meta-analysis. *J Neurol, 264*(4), 621-630.

Schuster, D. J., et al. (2014). Biodistribution of adeno-associated virus serotype 9 (AAV9) vector after intrathecal and intravenous delivery in mouse. *Front Neuroanat, 8*, 42.

Scriver, C. R. (1969). Treatment of inherited disease: realized and potential. *Med Clin North Am, 53*(4), 941-963.

Settembre, C., et al. (2013). Signals from the lysosome: a control centre for cellular clearance and energy metabolism. *Nat Rev Mol Cell Biol, 14*(5), 283-296.

Sevin, C., & Deiva, K. (2021). Clinical Trials for Gene Therapy in Lysosomal Diseases With CNS Involvement. *Frontiers in Molecular Biosciences, 8*.

Skaggs, B. J., et al. (2008). Induction of immune tolerance by activation of CD8(+) T suppressor/regulatory cells in lupus-prone mice. *Human Immunology, 69*(11), 790-796.

Sleat, D. E., et al. (2009). Mass spectrometry-based protein profiling to determine the cause of lysosomal storage diseases of unknown etiology. *Mol Cell Proteomics, 8*(7), 1708-1718.

Sly, W. S., et al. (1978). Receptor-mediated uptake of lysosomal enzymes. *Prog Clin Biol Res, 23*, 547-551.

Smallets, S., & Kendall, M. M. (2021). Post-transcriptional regulation in attaching and effacing pathogens: integration of environmental cues and the impact on gene expression and host interactions. *Current Opinion in Microbiology, 63*, 238-243.

Smid, B. E., et al. (2010). Pharmacological small molecules for the treatment of lysosomal storage disorders. *Expert Opin Investig Drugs, 19*(11), 1367-1379.

Smith, G. P. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science, 228*(4705), 1315-1317.

Smith, L. D., et al. (2020). Second Tier Molecular Genetic Testing in Newborn Screening for Pompe Disease: Landscape and Challenges. *Int J Neonatal Screen, 6*(2).

Sonawane, A. R., et al. (2017). Understanding Tissue-Specific Gene Regulation. *Cell Rep, 21*(4), 1077-1088.

Sonntag, F., et al. (2006). Adeno-associated virus type 2 capsids with externalized VP1/VP2 trafficking domains are generated prior to passage through the cytoplasm and are maintained until uncoating occurs in the nucleus. *J Virol, 80*(22), 11040-11054.

Sonntag, F., et al. (2010). A viral assembly factor promotes AAV2 capsid formation in the nucleolus. *Proc Natl Acad Sci U S A, 107*(22), 10220-10225.

Srivastava, A., et al. (1983). Nucleotide sequence and organization of the adeno-associated virus 2 genome. *J Virol, 45*(2), 555-564.

Staba, S. L., et al. (2004). Cord-blood transplants from unrelated donors in patients with Hurler's syndrome. *N Engl J Med, 350*(19), 1960-1969.

Staretz-Chacham, O., et al. (2009). Lysosomal storage disorders in the newborn. *Pediatrics, 123*(4), 1191-1207.

Stein, C. S., et al. (2001). In vivo treatment of hemophilia A and mucopolysaccharidosis type VII using nonprimate lentiviral vectors. *Mol Ther, 3*(6), 850-856.

Stok, M., et al. (2020). Lentiviral Hematopoietic Stem Cell Gene Therapy Corrects Murine Pompe Disease. *Mol Ther Methods Clin Dev, 17*, 1014-1025.

Strauss, T., & von Maltitz, M. J. (2017). Generalising Ward's Method for Use with Manhattan Distances. *PLoS One, 12*(1).

Suhonen, J., et al. (2006). Ex vivo and in vivo gene delivery to the brain. *Curr Protoc Hum Genet, Chapter 13*, Unit 13 13.

Sun, B., et al. (2005). Efficacy of an adeno-associated virus 8-pseudotyped vector in glycogen storage disease type II. *Mol Ther, 11*(1), 57-65.

Sweeney, N. P., & Vink, C. A. (2021). The impact of lentiviral vector genome size and producer cell genomic to gag-pol mRNA ratios on packaging efficiency and titre. *Molecular Therapy-Methods & Clinical Development, 21*, 574-584.

Takahashi, H., et al. (2002). Long-term systemic therapy of Fabry disease in a knockout mouse by adeno-associated virus-mediated muscle-directed gene transfer. *Proc Natl Acad Sci U S A, 99*(21), 13777-13782.

Tang, H., et al. (2020). The First Year Experience of Newborn Screening for Pompe Disease in California. *Int J Neonatal Screen, 6*(1), 9.

Tang, Q. Z., & Bluestone, J. A. (2008). The Foxp3(+) regulatory T cell: a jack of all trades, master of regulation. *Nature Immunology, 9*(3), 239-244.

Tarnopolsky, M., et al. (2016). Pompe Disease: Diagnosis and Management. Evidence-Based Guidelines from a Canadian Expert Panel. *Canadian Journal of Neurological Sciences, 43*(4), 472-485.

Terwilliger, E. F., et al. (1989). Construction and Use of a Replication-Competent Human Immunodeficiency Virus (Hiv-1) That Expresses the Chloramphenicol Acetyltransferase Enzyme. *Proceedings of the National Academy of Sciences of the United States of America, 86*(10), 3857-3861.

Teufel, F., et al. (2022). SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nature Biotechnology, 40*(7), 1023-+.

Thomas, E. D., et al. (1977). One hundred patients with acute leukemia treated by chemotherapy, total body irradiation, and allogeneic marrow transplantation. *Blood, 49*(4), 511-533.

Thomas, E. D., et al. (1957). Intravenous Infusion of Bone Marrow in Patients Receiving Radiation and Chemotherapy. *New England Journal of Medicine, 257*(11), 491-496.

Thorley, M., et al. (2016). Skeletal muscle characteristics are preserved in hTERT/cdk4 human myogenic cell lines. *Skeletal Muscle, 6*.

Todkar, K., et al. (2017). Mitochondria and Lysosomes: Discovering Bonds. *Front Cell Dev Biol, 5*, 106.

Toscano, A., & Schoser, B. (2013). Enzyme replacement therapy in late-onset Pompe disease: a systematic literature review. *J Neurol, 260*(4), 951-959.

Tran, M. Y., & Kamen, A. A. (2022). Production of Lentiviral Vectors Using a HEK-293 Producer Cell Line and Advanced Perfusion Processing. *Front Bioeng Biotechnol, 10*, 887716.

Trivedi, P. C., et al. (2020). Lysosomal Biology and Function: Modern View of Cellular Debris Bin. *Cells, 9*(5).

Tsoumpra, M. K., et al. (2019). Peptide-conjugate antisense based splice-correction for Duchenne muscular dystrophy and other neuromuscular diseases. *EBioMedicine, 45*, 630-645.

Uchida, N., et al. (2018). High-level embryonic globin production with efficient erythroid differentiation from a K562 erythroleukemia cell line. *Exp Hematol, 62*, 7-16 e11.

van den Hout, H., et al. (2000). Recombinant human alpha-glucosidase from rabbit milk in Pompe patients. *Lancet, 356*(9227), 397-398.

van den Hout, H. M., et al. (2003). The natural course of infantile Pompe's disease: 20 original cases compared with 133 cases from the literature. *Pediatrics, 112*(2), 332-340.

van der Beek, N. A., et al. (2012). Clinical features and predictors for disease natural progression in adults with Pompe disease: a nationwide prospective observational study. *Orphanet J Rare Dis, 7*, 88.

van der Horst, G. T., et al. (1987). Cell-free translation of human lysosomal alpha-glucosidase: evidence for reduced precursor synthesis in an adult patient with glycogenosis type II. *Biochim Biophys Acta, 910*(2), 123-129.

van der Ploeg, A. T., & Reuser, A. J. (2008). Pompe's disease. *Lancet, 372*(9646), 1342-1353.

van Gelder, C. M., et al. (2015). Enzyme therapy and immune response in relation to CRIM status: the Dutch experience in classic infantile Pompe disease. *J Inherit Metab Dis, 38*(2), 305-314.

van Meel, E., & Klumperman, J. (2008). Imaging and imagination: understanding the endo-lysosomal system. *Histochem Cell Biol, 129*(3), 253-266.

van Til, N. P., et al. (2010). Lentiviral gene therapy of murine hematopoietic stem cells ameliorates the Pompe disease phenotype. *Blood, 115*(26), 5329-5337.

Vauquelin, G., & Packeu, A. (2009). Ligands, their receptors and ... plasma membranes. *Mol Cell Endocrinol, 311*(1-2), 1-10.

Vellodi, A. (2005). Lysosomal storage disorders. *Br J Haematol, 128*(4), 413-431.

Vitner, E. B., et al. (2010). Common and uncommon pathogenic cascades in lysosomal storage diseases. *J Biol Chem, 285*(27), 20423-20427.

Vitner, E. B., et al. (2014). RIPK3 as a potential therapeutic target for Gaucher's disease. *Nat Med, 20*(2), 204-208.

Vogt, V. M., & Simon, M. N. (1999). Mass determination of rous sarcoma virus virions by scanning transmission electron microscopy. *J Virol, 73*(8), 7050-7055.

vom Dahl, S., & Mengel, E. (2010). Lysosomal storage diseases as differential diagnosis of hepatosplenomegaly. *Best Pract Res Clin Gastroenterol, 24*(5), 619-628.

Wada, M., et al. (2020). Ex Vivo Gene Therapy Treats Bone Complications of Mucopolysaccharidosis Type II Mouse Models through Bone Remodeling Reactivation. *Molecular Therapy-Methods & Clinical Development, 19*, 261-274.

Walterfang, M., et al. (2012). Dysphagia as a risk factor for mortality in Niemann-Pick disease type C: systematic literature review and evidence from studies with miglustat. *Orphanet J Rare Dis, 7*, 76.

Wang, D., & Gao, G. (2014). State-of-the-art human gene therapy: part II. Gene therapy strategies and clinical applications. *Discov Med, 18*(98), 151-161.

Wang, Y. C., et al. (2014). Protein post-translational modifications and regulation of pluripotency in human stem cells. *Cell Res, 24*(2), 143-160.

Watson, G. L., et al. (1998). Treatment of lysosomal storage disease in MPS VII mice using a recombinant adeno-associated virus. *Gene Ther, 5*(12), 1642-1649.

West, M., et al. (2009). Agalsidase alfa and kidney dysfunction in Fabry disease. *J Am Soc Nephrol, 20*(5), 1132-1139.

Wilusz, J. (2005). The Fundamentals of Human Virology. *Microbial Forensics*, 41-53.

Winchester, B. (2005). Lysosomal metabolism of glycoproteins. *Glycobiology, 15*(6), 1R-15R.

Winchester, B., et al. (2000). The molecular basis of lysosomal storage diseases and their treatment. *Biochem Soc Trans, 28*(2), 150-154.

Winkel, L. P., et al. (2004). Enzyme replacement therapy in late-onset Pompe's disease: a three-year follow-up. *Ann Neurol, 55*(4), 495-502.

Wolff, L., & Ruscetti, S. (1988). The spleen focus-forming virus (SFFV) envelope gene, when introduced into mice in the absence of other SFFV genes, induces acute erythroleukemia. *J Virol, 62*(6), 2158-2163.

Wong, L. F., et al. (2004). Transduction patterns of pseudotyped lentiviral vectors in the nervous system. *Mol Ther, 9*(1), 101-111.

Wong, L. F., et al. (2006). Lentivirus-mediated gene transfer to the central nervous system: therapeutic and research applications. *Hum Gene Ther, 17*(1), 1-9.

Wood, R. J., & Hulett, M. D. (2008). Cell surface-expressed cation-independent mannose 6-phosphate receptor (CD222) binds enzymatically active heparanase independently of mannose 6-phosphate to promote extracellular matrix degradation. *J Biol Chem, 283*(7), 4165-4176.

Woods, N. B., et al. (2003). Lentiviral vector transduction of NOD/SCID repopulating cells results in multiple vector integrations per transduced cell: risk of insertional mutagenesis. *Blood, 101*(4), 1284-1289.

Yaffe, D., & Saxel, O. (1977). Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. *Nature, 270*(5639), 725-727.

Yang, N. J., & Hinner, M. J. (2015). Getting across the cell membrane: an overview for small molecules, peptides, and proteins. *Methods Mol Biol, 1266*, 29-53.

Yin, H., et al. (2017). Delivery technologies for genome editing. *Nat Rev Drug Discov, 16*(6), 387-399.

Yoko-o, T., et al. (2021). Regulation of alcohol oxidase gene expression in methylotrophic yeast Ogataea minuta. *Journal of Bioscience and Bioengineering, 132*(5), 437-444.

Yoshida, T., et al. (2017). A Skeletal Muscle Model of Infantile-onset Pompe Disease with Patient-specific iPS Cells. *Sci Rep, 7*(1), 13473.

Zaiss, A. K., & Muruve, D. A. (2005). Immune responses to adeno-associated virus vectors. *Curr Gene Ther, 5*(3), 323-331.

Zani, I. A., et al. (2015). Scavenger Receptor Structure and Function in Health and Disease. *Cells, 4*(2), 178-201.

Zhang, X., et al. (2018). Blood-brain barrier shuttle peptides enhance AAV transduction in the brain after systemic administration. *Biomaterials, 176*, 71-83.

Zheng, C., & Baum, B. J. (2008). Evaluation of promoters for use in tissue-specific gene delivery. *Methods Mol Biol, 434*, 205-219.

Zhou, S., et al. (2016). Evaluating the Safety of Retroviral Vectors Based on Insertional Oncogene Activation and Blocked Differentiation in Cultured Thymocytes. *Mol Ther, 24*(6), 1090-1099.

Zincarelli, C., et al. (2008). Analysis of AAV serotypes 1-9 mediated gene expression and tropism in mice after systemic injection. *Mol Ther, 16*(6), 1073-1080.