PhD degree in Systems Medicine (curriculum in Human Genetics)

European School of Molecular Medicine (SEMM),

University of Milan and University of Naples “Federico II”

Settore disciplinare: MED/03

Advanced AAV-mediated liver-directed gene therapies for Haemophilia A and Mucopolysaccharidosis type VI

Federica Esposito

TIGEM, Pozzuoli
Matricola n. R13154

***Tutor**: Prof. Alberto Auricchio*

Scientific Director, Telethon Institute of Genetics and Medicine (TIGEM)

Dept. of Advanced Biomedical Sciences, University of Naples "Federico II"

***Internal Supervisor**: Prof. Nicola Brunetti-Pierri*

Dept. of Translational Medicine, University of Naples "Federico II"

Telethon Institute of Genetics and Medicine (TIGEM)

***External Supervisor**: Prof. Antonia Follenzi*

Dept. of Health Sciences, University of Piemonte Orientale "Amedeo Avogadro"

PhD Coordinator: Prof. Saverio Minucci

Anno accademico 2023-2024

Acknowledgments

At the end of this long journey, I want to express my gratitude to everyone who supported me and contributed to my success.

My heartfelt thanks go to my mentor, Prof.Alberto Auricchio, for the constant support and trust he has always placed in me. Thank you for guiding me while giving me the freedom to explore, make mistakes, and grow, but above all, for the countless opportunities you have provided. Without you, none of this would have been possible. I hope my work can reflect at least a part of the value of the lessons you've shared with me.

A special thanks goes to my entire group, to those who are with me today and to those who have been part of this journey in the past. A particular thanks to Arjun, Hristiana, Stefano, Mariangela, Elena, Miriam, Fabio, Paula, Patrizia, Emanuela, Roberto. Thank you for sharing with me your knowledge, experiences, challenges, and successes. Without your support and patience, the difficult moments would have been unbearable. Thank you for the respect and friendship you've shown me. You've made the dark days lighter and the successes even more meaningful. Each of you has left an indelible mark on this journey, and I will always be deeply grateful to you.

To my family, to whom I dedicate this achievement with love and pride, goes my deepest gratitude. In particular, thanks to my father and my mother, for always supporting me, even through the most difficult challenges. Without your support throughout all these years, I would never have been able to travel this path.

To all the people closest to me: Lucrezia, Francesca, Melania, Nina, Federica, Stefania, Anna, Luisa, Mery, Fatima, Marianna, Lucia, Viviana, Alfonso, Castiglia, Enzo, and especially to Roberto, who believed in me more than I ever did myself. Thank you for supporting me in moments of uncertainty and for reminding me each time that I had the ability to achieve my dreams.

It is thanks to all of you that I can celebrate this success today, which I dedicate with all my heart to each one of you.

Finally, thanks to myself, especially to my past self, for never giving up and for having the courage and tenacity to see it through to the end.

PUBLICATIONS

1) **Therapeutic homology-independent targeted integration in retina and liver.** Patrizia Tornabene, Rita Ferla, Manel Llado-Santaeularia, Miriam Centrulo, Margherita Dell'Anno, **Federica Esposito**, Elena Marrocco, Emanuela Pone, Renato Minopoli, Carolina Iodice, Edoardo Nusco, Settimio Rossi, Hristiana Lyubenova, Anna Manfredi, Lucio Di Filippo, Antonella Iuliano, Annalaura Torella, Giulio Piluso, Francesco Musacchia, Enrico Maria Surace, Davide Cacchiarelli, Vincenzo Nigro & Alberto Auricchio. Nature Communications volume 13, Article number: 1963 (2022). https://doi.org/10.1038/s41467-022-29550-8

2) **Liver gene therapy with intein-mediated F8 trans-splicing corrects mouse haemophilia A. Federica Esposito**, Hristiana Lyubenova, Patrizia Tornabene, Stefano Auricchio, Antonella Iuliano, Edoardo Nusco, Simone Merlin, Cristina Olgasi, Giorgia Manni, Marco Gargaro, Francesca Fallarino, Antonia Follenzi, and Alberto Auricchio. EMBO Mol Med (2022). DOI: 10.15252/emmm.202115199

3) **Full-length ATP7B reconstituted through protein trans-splicing corrects Wilson disease in mice.** Agnese Padula, Raffaella Petruzzelli, Sasha A Philbert, Stephanie J Church, **Federica Esposito**, Severo Campione, Marcello Monti, Filomena Capolongo, Claudia Perna, Edoardo Nusco, Hartmut H Schmidt, Alberto Auricchio, Garth J S Cooper, Roman Polishchuk, Pasquale Piccolo. Mol Ther Methods Clin Dev., (2022). DOI: 10.1016/j.omtm.2022.08.004

4) CoCas9 is a compact nuclease from the human microbiome for efficient and precise genome editing. Eleonora Pedrazzoli, Michele Demozzi, Elisabetta Visentin, Matteo Ciciani, Ilaria Bonuzzi, Laura Pezzè, Lorenzo Lucchetta, Giulia Maule, Simone Amistadi, **Federica Esposito**, Mariangela Lupo, Annarita Miccio, Alberto Auricchio, Antonio Casini, Nicola Segata & Anna Cereseto. Nature Communications, (2024). DOI: https://doi.org/10.1038/s41467-024-47800-9

5) Safe and effective liver-directed AAV-mediated homology-independent targeted integration in mouse models of inherited diseases. Federica Esposito, Fabio Dell'Aquila, Manuel Rhiel,Stefano Auricchio,Kay Ole Chmielewski, Geoffroy Andrieux, Rita Ferla, Paula Sureda Horrach, Arjun Padmanabhan, Roberto Di Cunto, Simone Notaro, Manel Llado Santeularia, Melanie Boerries, Margherita Dell'Anno, Edoardo Nusco, Agnese Padula ,Sofia Nutarelli, Tatjana I. Cornu, Nicolina Cristina Sorrentino, Pasquale Piccolo, Ivana Trapani, Toni Cathomen, Alberto Auricchio. Cell Reports Medicine, (2024) https://doi.org/10.1016/j.xcrm.2024.101619

INDEX

ABBREVIATIONS

AAP: assembly-activating protein

AAV: Adeno-associated virus

ARSB: arylsulfatase B

BDD: B-domain deleted

bGH: bovine growth hormone polyA

bp: base pairs

CMV: cytomegalovirus promoter

Cbh: Chicken beta-actin hybrid promoter

CDS: coding sequence

CRISPR: clustered regularly interspaced short palindromic repeats

DEL: deletion

DNA: deoxyribonucleic acid

DSB: double-strand break

DsRed: Dicosoma red fluorescent protein

EDTA: etilendiaminotetraacetic acid

EGFP: Enhanced green fluorescent protein

ERT: enzyme replacement therapy

F8: Factor F8 gene

FACS: fluorescence activated cell sorting

Fwd: forward

GAGs: glycosaminoglycans

GC: genome copies

GE: Genome Editing

gRNA: guide RNA

GT: Gene Therapy

HEK293: human embryonic kidney 293

HDR: homology-directed repair

HITI:homology-independent targeted integration

HLP: hybrid liver promoter

IEM: Inborn errors of metabolism

INDELs: insertions and deletion

INS: insertion

IPSC: induced pluripotent stem cells

ITR: Inverted Terminal Repeats

Kb: Kilo bases

KO: knockout

LSD: lysosomal storage disease

MPS: Mucopolysaccharidosis

MPSVI:Mucopolysaccharidosis type VI

mRNA: messenger RNA

NGS: next-generation sequencing

NHEJ: non-homologous end joining

ORF: open reading frame

PAM: protospacer-adjacent motif

PCR: polymerase chain reaction

PFA: paraformaldehyde

polyA: polyadenylation signal

PTS: protein *trans*-splicing

Rev: reverse

RNA: Ribonucleic acid

RT-PCR: real-time PCR.

SAS: Synthetic splicing acceptor signal

shpolyA: short synthetic polyA

SpCas9: Steptococcus pyogenes Cas9

TBG: thyroxin-binding globulin promoter

TIDE: Tracking of INDELs by Decomposition

tracrRNA: trans-activating crRNA

T2A: Thosea asigna virus 2A skipping peptide

T7E: T7 Endonuclease I

UTR: untranslated region

Vg: viral genome

WT: wild type

FIGURES AND TABLE INDEX

ABSTRACT

Liver-directed gene therapy using adeno-associated viral (AAV) vectors holds significant promise for providing long-term transgene expression following a single systhemic administration. However, both the limited cargo capacity (~4.7 kb) and the non-integrative nature of AAV vectors prevent their application to conditions caused by either mutations in large genes (i.e. *F8* whose coding sequence exceeds ~ 4.7 kb) or to those that require early stage of intervention, such as early-onset inborn errors of metabolism. My PhD project aims to overcome these shortcomings and broaden the therapeutic applications of AAVs. To overcome the issue of limited cargo capacity, I explored an AAV-mediated protein trans-splicing (PTS) strategy utilizing split inteins. PTS is a natural mechanism employed by organisms across all domains of life to reconstitute large proteins from shorter precursors. Split inteins facilitate this process by self-excising from the host protein in an energy-independent manner. I used this approach for haemophilia A (HemA), the most common X-linked bleeding disorder (affecting 1 in 5,000 males), caused by a deficiency of clotting factor 8 (*F8*, ~7 kb). I first demonstrated the successful full-length reconstitution of the large (~5 kb) and highly active B-domain deleted (B-DD) N6-*F8* variant in vitro. Subsequently, I designed two separate AAV vectors, with regular size genomes, each encoding for half of the N6-*F8***,** flanked by split inteins. The AAV-N6 intein system led to successful reconstitution of functional full-length N6-F8 in the mouse liver, achieving therapeutic F8 activity in haemophilic mice without eliciting anti-F8 immune responses.

In parallel, I developed an AAV-mediated CRISPR/Cas9 homology-independent targeted integration (AAV-HITI) platform to achieve stable transgene integration at the 3' end of the mouse Albumin locus, which is highly expressed in hepatocytes. This approach mitigates the loss of episomal AAV genomes in proliferative tissues, thereby preserving therapeutic efficacy over time.

I demonstrated AAV-HITI efficacy, in mouse models of HemA and Mucopolysaccharidosis type VI (MPS VI), a lysosomal storage disorder, achieving stable therapeutic levels of systemic proteins even at low AAV doses. Importantly, I performed comprehensive molecular analyses to characterize the AAV-HITI outcomes within the target site confirming successful targeted integration. Moreover, in AAV-HITI treated livers nor gross chromosomal rearrangements or significant off-target insertions/deletions were observed 1-year post neonatal delivery. Furthermore, I demonstrated the effectiveness of this platorm in adult MPS VI and HemA mice, using vector doses that are within the safe range for human translation.

Overall, my findings support two distinct AAV-mediated liver-directed platforms to address major limitations of current AAV-based therapies. The AAV-intein N6-F8 system holds promise as a potential therapeutic option for HemA, while the AAV-HITI platform could serve as a durable therapeutic option for liver-targeted diseases, particularly those requiring early intervention.

These innovations represent significant advancements in overcoming challenges related to cargo capacity, stable transgene expression, and long-term efficacy in gene therapy.

INTRODUCTION

Haemophilia A

With an incidence of 1 in 5,000 male live births haemophilia A (HemA, OMIM#306700) is the most common X-linked recessive coagulation disorder caused by a deficiency or complete absence of coagulation factor VIII (F8)[1,2] a cofactor involved in the intrinsic pathway of the coagulation cascade. Clinical severity is determined by the extent of residual F8 activity and can be classified as mild (5-50% F8 activity), moderate (1-5% F8 activity) or severe (<1% F8 activity). The severe HemA is characterised by spontaneous musculoskeletal and soft tissue bleeding as well as the inability to achieve haemostasis after trauma unless concentrates of clotting factor are infused [3].

Current treatments for haemophilia A

Treatments for severe haemophilia A involve the use of F8 replacement or an antibody-based (Emicizumab) therapies to improve the ability of blood to clot and reduce the likelihood of bleeding. However, periodic injections are still required, thus requiring constant monitoring of treatment effectiveness and safety [4,5,6]. In the last decades the use of gene therapy for HemA has been under extensive investigation[7] given the monogenic nature of the disease and considering that modest improvements in F8 levels (by 1-2%) can ameliorate the haemophilic phenotype reducing the risk of spontaneous bleeding events and the need for infusions. Several gene transfer strategies for F8 replacement have been evaluated, in combination with adeno-associated viral (AAV) vectors[8,9]. However, AAV cargo capacity (~4.7 kb) is exceeded by the F8 coding sequence (CDS;~7 kb), which poses a significant challenge to AAV gene therapy. Since B-domain of F8 is dispensable for its normal coagulation activity [10,11], current pre-clinical and clinical gene therapy approaches rely on B-domain deleted (BDD) and codon-optimized F8 forms, utilizing both short liver-specific promoters and

poly-adenylation (PolyA) signals [12,13]. Valoctocogene roxaparvovec (marketed as Roctavian), is an AAV5 vector delivering a B-domain-deleted *F8* transgene that has been conditionally approved in Europe in August 2022 and approved in the United States in June 2023[6]. Clinical studies have shown that patients treated with a single Roctavian intravenous infusion, exhibited expression of F8 and significant reductions in annualized bleeding rates. Despite this great advancement in the field, challenges such as long-term efficacy and potential immune responses continue to be areas of active research[6].

Inborn errors of metabolism

Inborn errors of metabolism (IEM) are a diverse group of genetic disorders usually caused by deficiency of an enzyme needed to convert one metabolite to another resulting in the accumulation of a metabolic intermediate that are normally present in low concentrations. Metabolic intermediates are usually not toxic, but high levels can cause serious effect. IEM manifest with a broad spectrum of clinical symptoms depending on the affected metabolic pathway. Early diagnosis through expanded newborn screening has significantly improved the prognosis of many IEMs, enabling timely interventions to prevent severe complications[14]. Advances in genomic technologies, such as next-generation sequencing, have facilitated the identification of novel mutations and their pathogenic mechanisms which underscores the importance of genetic counseling and personalized medicine[15]. Current treatment are based on enzyme replacement therapies (ERT), particularly for lysosomal storage disorders, and have shown promising results, although challenges remain regarding the delivery and immune responses and frequent treatment [16,17]. Moreover, novel therapeutic approaches, including gene therapy and small molecule chaperones, are being explored, offering potential therapeutic options for conditions previously deemed untreatable.

Lysosomal storage diseases

Lysosomal storage diseases (LSDs) are a group of IEMs caused by defects in lysosomal hydrolases, resulting in substrate accumulation and cellular dysfunction [18]. The lysosome is a membrane-bound organelle, crucial for cellular digestion and waste processing. It contains hydrolytic enzymes that break down proteins, lipids, nucleic acids, and carbohydrates in an acidic environment (pH 5; [19]. Lysosomal hydrolases are targeted to the lysosome via mannose 6-phosphate, which is recognized by lysosomal membrane receptor. Importantly, lysosomal enzymes can also be secreted and then uptaken by other cells via mannose 6-phosphate receptors in the plasma membrane, followed by trafficking to the lysosome. Internalized molecules are then degraded within lysosomes while cell's components are recycled. This turnover is essential for cellular homeostasis and stress response[20]. Mucopolysaccharidoses (MPS) are a group of LSDs characterized by the accumulation of glycosaminoglycans (GAGs) due to deficiencies in specific lysosomal enzymes responsible for their degradation. This accumulation leads to progressive cellular, tissue, and organ damage, manifesting in symptoms such as skeletal deformities, organomegaly, and neurological impairment[16]. Mutations in different genes cause accumulation of different GAGs, which has different effects, giving raise to 7 main kinds of MPS (MPSI-MPSVII).

Mucopolysaccharidosis type VI

Mucopolysaccharidosis type VI (MPS VI) also known as Maroteaux-Lamy Syndrome is a lysosomal storage disorder which affects 1-9/1.000.000 people worldwide (https://www.orpha.net/) and is caused by mutations in the Arylsulfatase B gene (ARSB) located in chromosome 5[21]. ARSB is a lysosomal enzyme involved in the hydrolysis of the C4-sulfate ester link of dermatan sulfate during its lysosomal processing. ARSB deficiency leads to abnormal storage in different organs such as: liver, kidney, spleen, heart with no impairment of the central nervous systhem. Of note, the measurement of GAGs levels in the urine samples is considered a useful biomarker for MPS VI diagnosis [22].

Current treatments for MPS VI

The standard of care for several LSDs, including MPSVI is Enzyme Replacement Therapy (ERT). This treatment consists in weekly intravenous administrations of an exogenous recombinant human ARSB (galsulfase) which results in reduction of GAGs in affected tissues and urine. However, it requires life-long and costly intravenous infusions and has limited efficacy on skeletal abnormalities, cardiac valves, and corneal opacities [22]. Considering the ERT limitations, new therapeutic approaches are needed. Stem cell transplantation (HSCT) has showed therapeutic outcomes similar to ERT however it has also faced important drawbacks like graft failure, limited donor availability, and risks from preconditioning and graft-versus-host disease[23]. In the last years in-vivo liver-directed gene therapy, particularly with AAV has emerged as a powerful strategy to treat MPSVI as well as other LSDs [21,22,24–28]. However, the non-integrative nature of AAV genomes, in some instances, prevents their use from neonatal treatments [29–32] which is the preferred stage of intervention for diseases such as early-onset inborn errors of metabolism.

Gene therapy

The core principle behind gene therapy (GT) is to correct inherited disorders by delivering exogenous therapeutic genes into target cells, using a suitable vector [33]. Moreover, recent advancements have expanded the potential of gene therapy including the use of CRISPR-Cas9 technology for precise genome editing (GE). This approach allows for the direct modification of specific DNA sequences within the genome [34]. GT can be performed in different ways: in vivo, directly to the target tissues (animal model or patient) or ex vivo, in cultured cells isolated from animal or patient and re-implanted to it after treatment (**Figure 1**;Goswami *et al*, 2019). Of note, GT clinical applications are vast, ranging from the treatment of monogenic disorders, such as haemophilia A or B, to complex conditions like cancer and neurodegenerative diseases [36]. In all cases, the crucial factor for successful GT is an efficient delivery to the target cells. Ideal GT vectors should have the following features: i) the ability to localize to target cells and release their genetic material inside the nucleus; ii) provide long-term expression of the delivered transgene; iii) have sufficient DNA capacity to package the required genetic material; iv) elicit low to no toxicity. Powerful GT vectors are manly divided in two different classes: viral-derived vectors and non-viral vectors [37]. Viral-derived vectors in which viral replication is eliminated, exploit the ability of viral capsids to infect cells and transfer DNA to their nucleus. GT viral vectors, derive from different viruses such as lentiviruses, adenoviruses, and adeno-associated viruses (AAV). Lentiviral vectors, originate from human immunodeficiency virus 1 (HIV-1) and can stably integrate into the host genome infecting both dividing and non-dividing cells, making them suitable for long-term gene expression [38]. Adenoviral vectors are known for their large transgene capacity and strong transient expression but can elicit significant immune responses [39]. AAV vectors exhibit low immunogenicity and excellent ability to provide long-term expression in non-dividing cells; morover, these vectors have

shown efficacy in treating different genetic disorders and have been approved for several clinical applications[40]. Non-viral vectors represent a promising alternative in GT due to their reduced immunogenicity and ease of production compared to viral vectors. These vectors include liposomes, polymer-based systems and plasmid DNA, which can be engineered to deliver therapeutic genes into target cells. Recently, novel delivery systems like lipid nanoparticles and cationic polymers, which enhance cellular uptake and transgene expression, are under evaluation to improve the efficiency of these non-viral vectors [37]. Additionally, the use of physical methods such as electroporation and ultrasound-mediated delivery has shown potential in overcoming cellular barriers and facilitating the entry of non-viral vectors into cells [41]. Each of these vector offers unique advantages and challenges, and ongoing research aims to enhance their safety, specificity, and therapeutic potential in gene therapy.

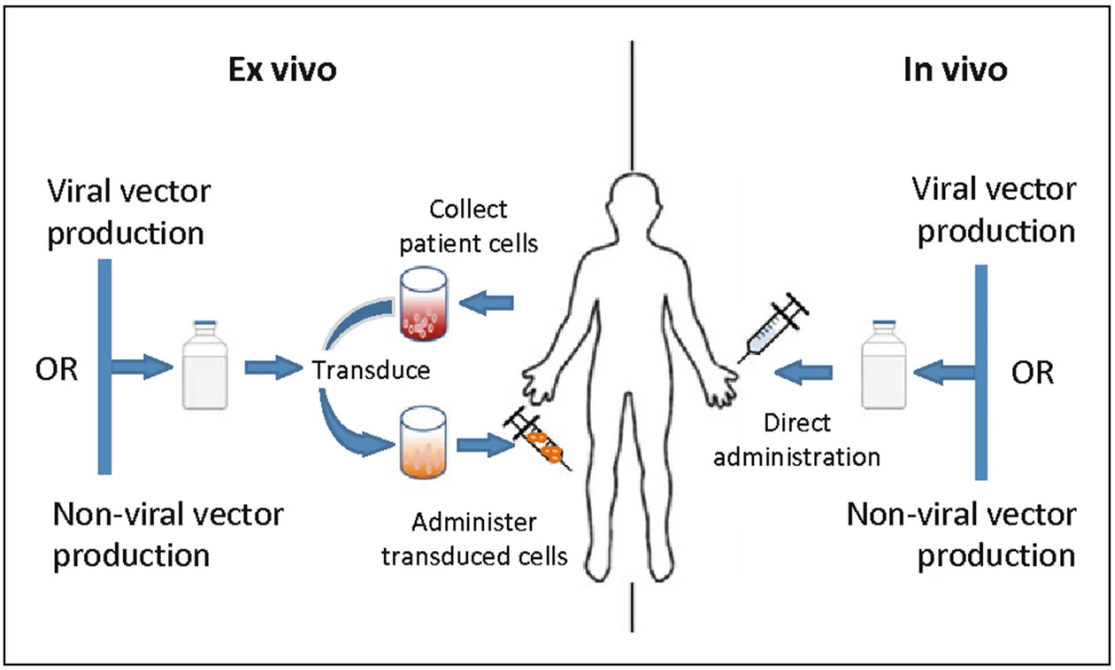


Figure 1. Strategies of ex vivo and in vivo gene therapy. *Adapted from* [42].

Adeno associated viral (AAV) vectors

Adeno-associated viruses (AAVs) are part of the Parvoviridae family and are classified as dependoviruses because they require co-infection with a helper virus for replication. AAVs have a single-stranded DNA genome of approximately 4.7 kilobases (kb) encapsidated within a non-enveloped icosahedral capsid. The genome features two inverted terminal repeats (ITRs) of 145 nucleotides each and two open reading frames (ORFs) encoding the Rep and Cap genes [43]. The Rep gene encodes for proteins necessary for replication, and packaging of newly synthesized single stranded AAV genomes. Rep also induces AAV genome integration in the AAVS1 locus on human chromosome 19. The Cap gene encodes the capsid proteins: VP1, VP2, and VP3, which assemble into the viral capsid. The AAV capsid, approximately 25 nm in diameter, is composed of 60 subunits of VP1, VP2, and VP3 in a roughly 1:1:10 ratio respectievely. Moreover, within the Cap gene alternative ORFs have been recently described. A first alternative ORF econdes the assembly activating protein (AAP). AAP targets newly synthesized capsid proteins to the host cell nucleus, favouring the formation of new virions [44] **Figure 2**). Another alternative ORF encodes a membrane-associated accessory protein (MAAP). MAAP contains a highly conserved, cationic amphipathic domain critical for AAV secretion [45].

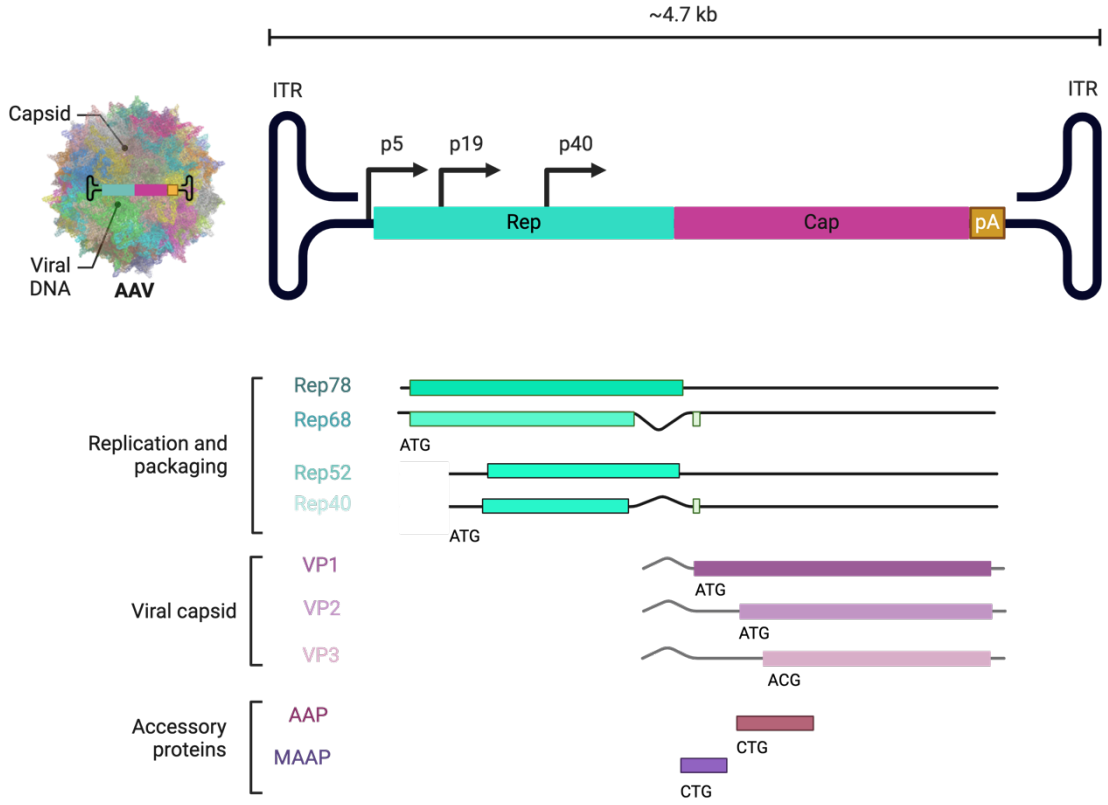


Figure 2. Adeno-associated viral (AAV) genome.

AAV genome is a linear single stranded DNA of 4.7 kb. The genome includes two different ORF flanked by two sequences (145bp) of intevred terminal repeats. One ORF encodes the Rep proteins: Rep78, Rep68, Rep52, Rep 40 required for viral replication and virion assembly. The second ORF encodes the viral capsid proteins named VP1,VP2,VP3. Alternative ORFs encode accessory proteins AAP and MAAP.

AAV capsid plays crucial roles in the entry into target cells, making it a key determinant of the vector's tropism. The viral infection starts when AAVs bind to different cell surface receptors and co-receptors, depending on the capsid serotype. For instance, AAV2 primary receptor is heparan sulfate proteoglycan and fibroblast growth factor receptor 1 (FGFR1) or αVβ5 integrin as co-receptors [46]. Following receptor binding, the virus is internalized via clathrin-mediated endocytosis [47,48]. Post-internalization, AAVs traverse the endosomal pathway, where endosomal acidification triggers capsid conformational changes which facilitate its escape into the cytoplasm [48]. The virus then moves toward the

nucleus, likely utilizing the cell's microtubule network for transport. Upon reaching the nuclear envelope, AAV uncoats and releases its single-stranded DNA genome into the nucleus, where it is converted to double-stranded DNA ultimately resulting in transcription and expression of the viral genome (**Figure 3**)[49].

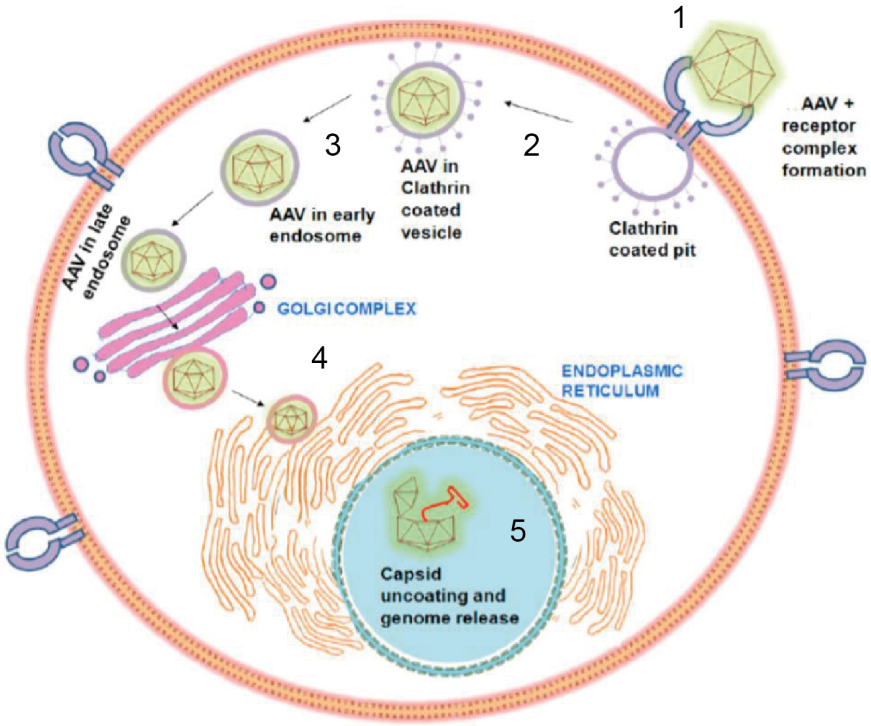


Figure 3. Adeno-associated viral infection.1**)** AAV particles bind to the cell surface receptor to form an AAV-receptor complex. 2) This complex is then internalized via endocytosis. 3) Intra-cellular trafficking is then initiated with the formation of endosomes. A typically low pH of 6 in early endosome and of 5 in late endosome induces a conformational change in the AAV capsid to expose the N-terminal phospholipase A2 domain and nuclear localization signal. 4) This facilitates AAV endosomal escape and then AAV particles traffic through both the Golgi complex and endoplasmic reticulum via the retrograde transport mechanism. 5) Upon translocation into the nucleus, the virus undergoes uncoating of its capsid and the genome is released. The genome might then integrate into the host genome at the AAVS1 site [50].

Differently from wild-type AAV genomes which may undergo integration into the human chromosome 19 Rian locus [51] recombinant AAVs (rAAVs) lack the rep gene and have a reduced likelihood of integrating into the genome and generally remain episomal. In rAAVs the two ITRs sequences are the minimal cis-acting elements that are required for the encapsidation of the transgene. This allows the substitution of 95% of the original genome with a DNA sequence of interest (**Figure 4**).

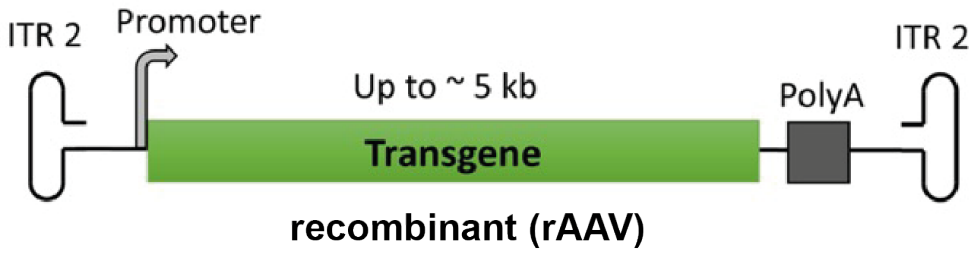


Figure 4. Recombinant AAV vectors.The rAAV viral genome consists of a transgene expression cassette contained between the ITRs sequences from the AAV serotype 2.

For viral preparation the rep and cap sequences are provided in trans together with the adenoviral helper genes [50]. The triple-transfection method is the most commonly used for the production of rAAV vectors at high yields. It is based on the co-transfection of permissive cells, usually human embryonic kidney 293 cells (HEK293), with three plasmids: i) a plasmid containing the transgene of interest flanked by the ITRs usually from AAV serotype 2, commonly referred as "transgene expression cassette", ii) a packaging plasmid containing the rep and cap genes, and iii) a plasmid encoding for adenoviral helper genes (**Figure 5**). Column chromatography or Cesius Cloride (CsCl) gradient centrifugation are then used to purify rAAV from both cellular contaminants and empty viral capsids [52].

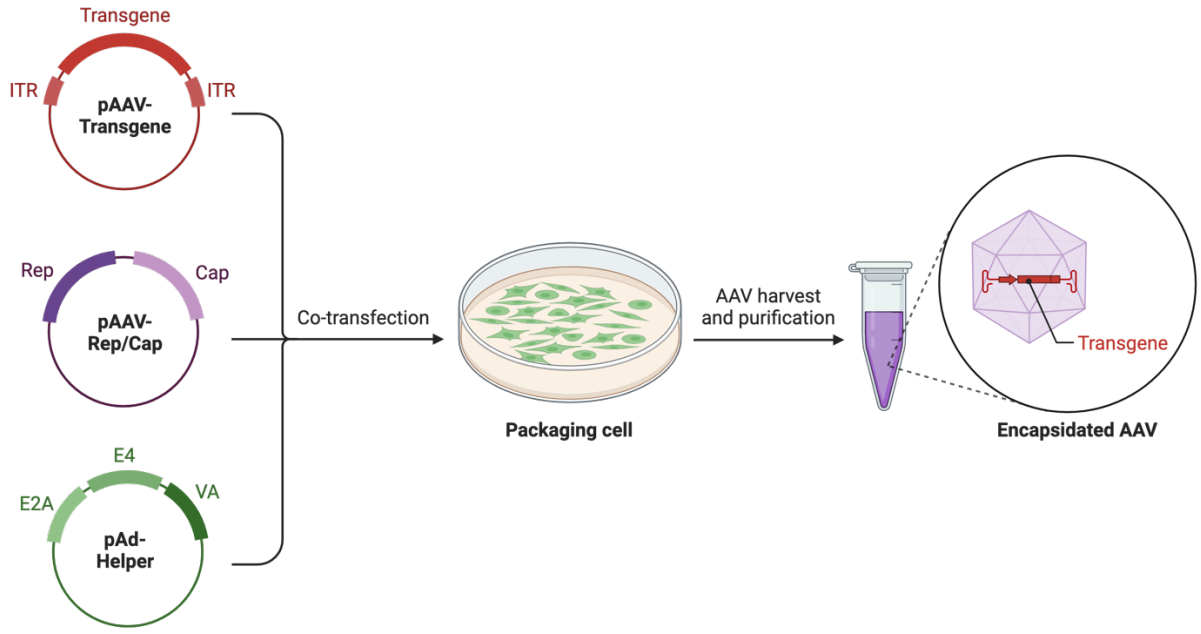


Figure 5. rAAV production by triple transfection. In the triple transfection method, HEK293 cells which endogenously express adenovirus E1a and E1b are cotransfected with: a plasmid carrying the transgene expression cassette between the AAV2 ITRs (pAAV-Transgene); a plasmid containing the AAV rep/cap sequences (pAAV Rep/Cap); a plasmid containing the remaining adenovirus helper genes (pAd-Helper).

AAV-liver directed gene therapy

The liver is a critical target tissue for gene therapy due to its central role in metabolism, detoxification, and synthesis of essential proteins. As the body's metabolic hub, the liver is responsible for storage and release of glucose, for breakdown of toxins and waste products, [53] and for the production of blood clotting factors and enzymes involved in metabolic pathways. This makes the liver an actractive target for therapies aimed at correcting genetic deficiencies or metabolic disorders. Liver-directed gene therapy with recombinat adeno-associated viral (rAAV) vectors has gained considerable attention as a transformative approach for treating genetic and acquired liver disease. Following a single intravenous administration, rAAV vectors are transported through the bloodstream to the liver where they primarily transduce hepatocytes. Once inside the targeted cells, rAAV vectors release their genetic material, which is then transcribed and translated into the therapeutic protein [54]. This process converts hepatocytes into factories for efficient and sustained transgene expression, enabling the secretion of therapeutic proteins into the bloodstream (**Figure 6**) and providing long-term therapeutic effects in both preclinical studies and humans [22,55,56] .

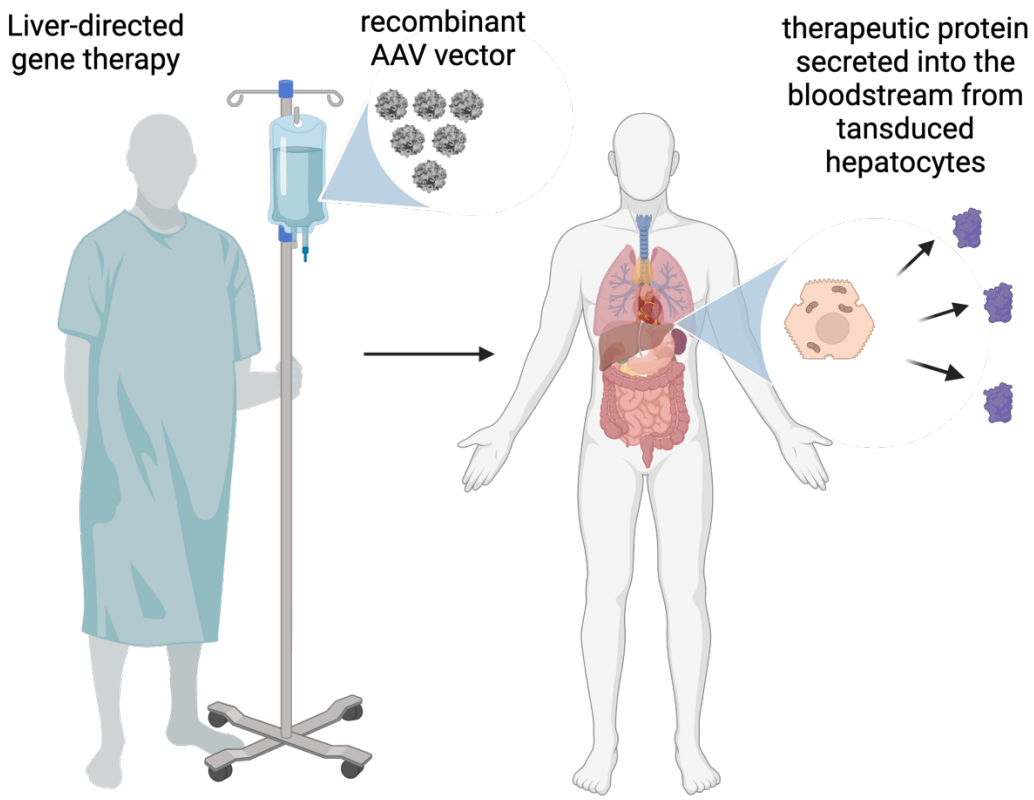


Figure 6. rAAV-mediated liver-directed gene therapy. Following a single intravenous rAAV administration the liver is converted into a factory for systemic secretion of therapeutic proteins into the bloodstream.

Recent advancements and clinical successes, including the approval of novel gene therapies by the U.S. Food and Drug Administration (FDA), highlight the potential of this approach. In 2022, the FDA approved etranacogene dezaparvovec (marketed as Hemgenix), an AAV5 vector encoding the Padua variant of factor IX (FIX) for the treatment of haemophilia B[3-57]. Similarly, gene therapy for haemophilia A has seen promising developments. In 2023, the FDA approved valoctocogene roxaparvovec (marketed as Roctavian), a rAAV5 delivering a B-domain-deleted F8 variant [58–60]. These advancements underscore the liver's significance as a target for gene therapy, paving the way for novel treatments for a variety of genetic and metabolic disorders.

Limits of liver-directed gene therapy with rAAVs

Despite the promising outcomes, several challenges remain in AAV-mediated liver-directed gene therapy. One of the main limitations is the immune response against AAVs capsid; elevated alanine aminotransferase (ALT) levels that indicate liver inflammation, have been observed upon AAV administration [58]. These responses can diminish the efficacy of gene therapy and necessitate the use of immunosuppressive regimens. Ongoing research aims to refine these regimens to balance immune suppression and therapeutic efficacy[58]. Moreover, the generation of neutralizing antibodies following the initial AAV administration prevents repeated dosing with the same vector serotype [61]. A further important limitation is constituted by the rescricted AAV vector cargo capacity (4.9Kb)[62,63]. This constrains the size of the therapeutic gene and regulatory elements that can be delivered, posing challenges for treating diseases that require large genes (i.e. F8 CDS ∽7 Kb). Moreover, the non-integrative nature of AAV genomes excludes the use of these vectors from neonatal treatments, which are the preferred stage of intervention for diseases such as early-onset inborn errors of metabolism [29–31]. Liver growth in young subjects leads to AAV genome loss [64] over time resulting in the decline of the therapeutic efficacy following cell division [22-65]. Despite these limitations, AAV-based gene therapy looks promising, and several research focusing on enhancing vector design and delivery methods are ongoing to overcome these shortcoming.

Novel AAV capsid development

The development of novel AAV capsids with improved liver tropism and reduced immunogenicity is a crucial area of interest in gene therapy research. Enhancing liver tropism involves engineering AAV capsids to more effectively target hepatocytes, thereby improving therapeutic outcomes for liver-associated diseases. Recent studies have shown that introducing liver-specific peptide motifs on the capsid surface can significantly enhance hepatocyte transduction efficiency [66,67]. Additionally, the use of high-throughput screening and directed

evolution techniques has allowed for the identification of capsid variants that show improved liver targeting and reduced off-target effects[68]. Reducing immunogenicity is another critical objective, as pre-existing antibodies against AAV capsids can neutralize the vectors and reduce therapeutic efficacy. Recent research has focused on rational design and synthetic biology approaches to create capsid variants that can evade the host immune response. For example, the use of synthetic capsid libraries has enabled the selection of AAV variants with mutations that reduce the binding to neutralizing antibodies while maintaining high transduction efficiency[69,70]. Furthermore, combining these engineered capsids with immune modulation strategies, such as transient immunosuppression, has shown promise in further reducing immunogenicity and allowing for repeated dosing [71,72].Advancements in capsid engineering are critical for expanding the therapeutic applications of AAVs, particularly for chronic liver diseases, and improving the overall safety and efficacy of AAV-mediated liver-directed gene therapy.

Protein *trans*-splicing

One innovative strategy to overcome the limited AAV cargo capacity is the intein-mediated protein trans-splicing (PTS) mechanism. This process involves the use of elements called intein that have the intrinsic ability to promote protein splicing. Intein-mediated PTS involves the excision of the intein from a precursor protein and the ligation, through a peptide bond, of the remaining external protein fragments (extein), resulting in a mature and functional protein. Inteins have been identified across various domains of life, including bacteria, archaea, and eukaryotes [73,74]. Intein-mediated PTS is an autocatalytic process that does not need external cofactors, but require a precise and highly conserved sequence of chemical reactions. These steps involve the formation of a thioester or ester intermediate, transesterification, and a peptide bond rearrangement, ultimately excising the intein and joining the exteins [75,76]. Most inteins are expressed within a single polypeptide chain (cis-splicing inteins; **Figure 7A**) but some are split into two polypeptides (trans-splicing inteins; **Figure 7B**).

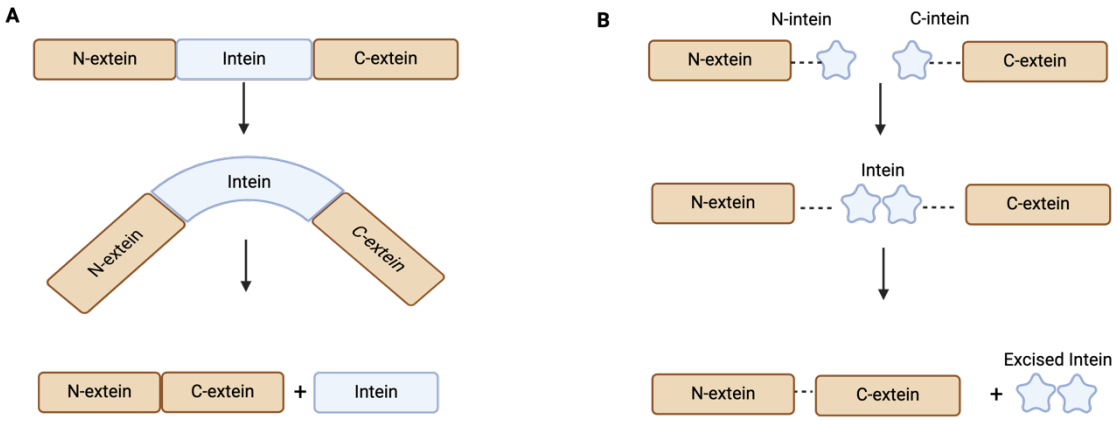


Figure 7. Protein splicing of Inteins.Schematic of Intein cis-splicing is represented in **A** and split-intein trans-splicing is represented in **B**. Adapted from [77].

Protein trans-splicing, i.e. fusion and splicing between two different proteins, named N and C fragments, is mediated by split-inteins of about 30 to 100 amino acids in length fused at the C- and N-terminal of the N and C fragments, respectively. The first naturally occurring split-intein was identified in DnaE from cyanobacterium Synechocystis sp. strain PPC6803 [78]. The N- and C-terminal halves of DnaE are encoded by two separate genes: one gene product consists of the N-terminal half of DnaE followed by 123 amino acids (a.a.) intein fragment, the second gene consist of 36 a.a. intein fragment followed by the C-terminal half of DnaE. When co-expressed, the two protein products associate rapidly to form the intact DnaE protein [79]. The unique properties of inteins have been harnessed for numerous applications and have been pivotal in the field of synthetic biology and protein engineering. Inteins have been exploited for protein ligation approaches enabling the site-specific conjugation of proteins, peptides, or other molecules.

In the gene therapy field, split-inteins offer innovative solutions to overcome the size limitations of adeno-associated viral (AAVs) vectors [80,81]. By splitting a large therapeutic transgene into smaller, vector-compatible fragments, split-inteins are able of catalyzing their own excision while promoting the ligation of flanking elements reconstituting the full-length protein in vivo (**Figure 8**) [81–83]. This approach has been particularly beneficial in delivering large genes, such as dystrophin for the treatment of Duchenne muscular dystrophy, thereby broadening the therapeutic potential of gene therapies [84].

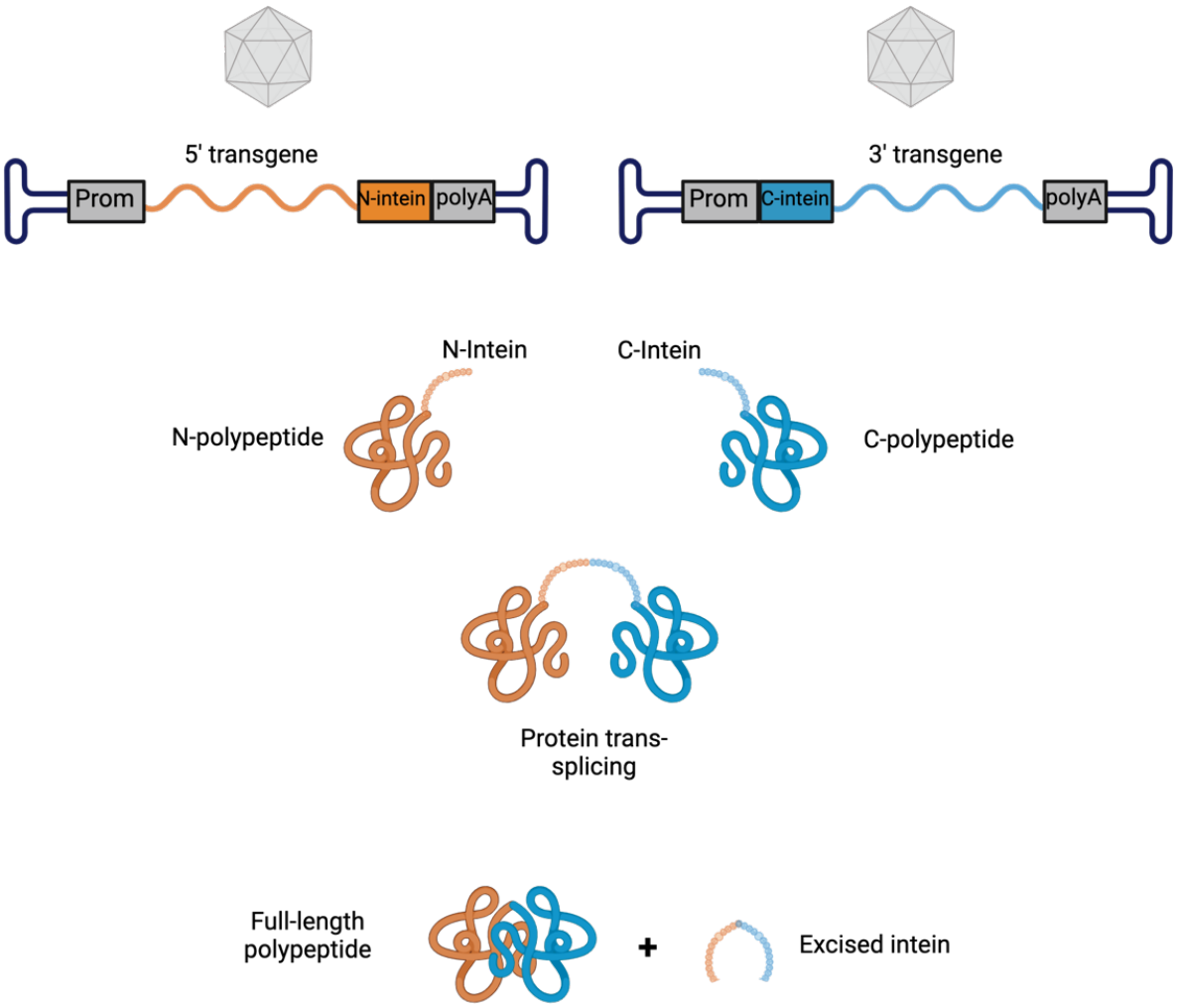


Figure 8. AAV-split-intein mediated PTS. Schematic of the split-intein mediated PTS adapted with AAV. Prom: promoter; 5'transgene: 5' half of the transgene CDS; N-intein: N-terminal of the intein;polyA: polyadenilation singal; 3'transgene: 3' half of the transgene CDS; C-intein: C-terminal of the intein.

Recent advancements in intein research have focused on enhancing their efficiency, specificity, and versatility. Engineered inteins with faster splicing kinetics have been developed, improving their utility in various applications. Moreover, intein-based technologies are being integrated with other cutting-edge biotechnological tools, such as CRISPR-Cas systems, to enable precise GE and functional protein modifications [85,86]. In conclusion, intein-mediated PTS represents a significant advancement in overcoming the size limitations of AAV

vectors, thus expanding the potential of AAV-mediated gene therapy to treat a wider range of genetic disorders and complex diseases.

CRISPR/Cas9

The CRISPR/Cas9 system has revolutionized genetic engineering, providing a powerful tool for precise GE across various organisms. CRISPR/Cas9 originally derived from the adaptive immune system of bacteria is based on the ability to cut small DNA fragments from pathogens storing them in a DNA array called Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR; **Figure 9A**) [87,88]. The CRISPR array is then transcribed into RNA molecules called crRNA, each corresponding to a different foreign DNA fragment. By a trans-activating crRNA (tracrRNA) the different crRNA bind to the Cas9 endonuclease forming a ribonucleoprotein complex (**Figure 9B**).

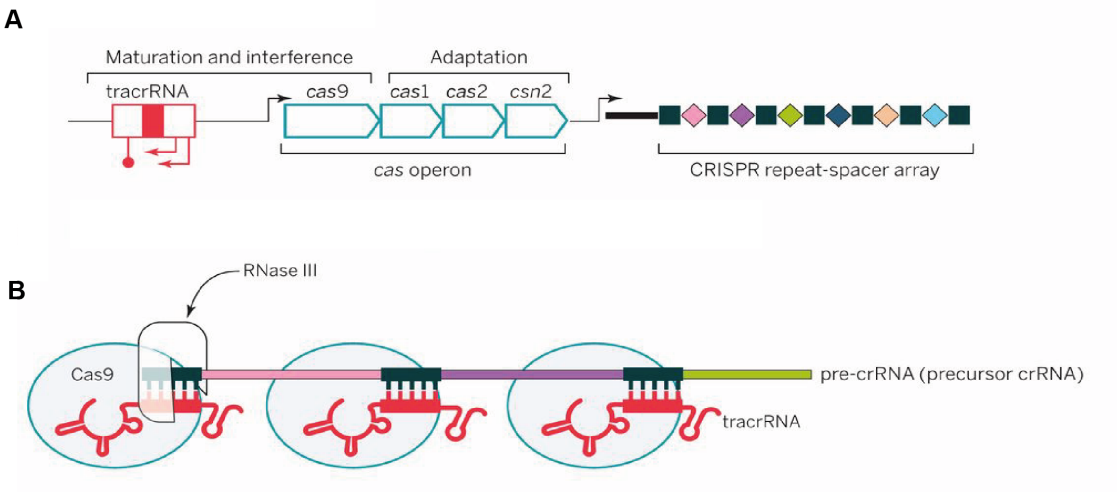


Figure 9. Shematic of the type II-A system from S. pyogenes. A) The cas gene operon with tracrRNA and the CRISPR array is shown. **B**) tracrRNA:crRNA duplex and Cas9 co-complex formation. Cas9 association with the antirepeat-repeat RNA (tracrRNA:crRNA) duplexes, RNA co-processing by ribonuclease III, further trimming, R-loop formation, and target DNA cleavage. Adapted from Doudna and Charpentier [34].

Cas9 consists of five different domains: RuvC domain, BH domain, REC domain, HNH domain, and PI domain, which are divided into nuclease (NUC) domains and recognition (REC) domains which undergoes significant conformational changes upon binding of tracrRNA and crRNA to its RecI domain[89]. The ribunocleoprotein compex can recognize any DNA sequence complementary to the crRNA, as well as a 3-5 bp sequence called protospacer-adjacent motif (PAM) at the 3' end of the recogition site which is not part of the crRNA and is recognized by the PAM-interacting domain of the Cas9 protein. The absence of the PAM sequence within the CRISPR array in the bacterial genome avoids its cleavage by Cas9[89]. Once the Cas9 ribonucleoprotein recognizes the target sequence, the HNH and RuvC domains cleave the target and non-target strands respectively, generating a double-strand break (DSB; **Figure 10A**) [34,90]. This break is then repaired by the cellular mechanisms such as non-holomogous end joingin or homologous recombination, leading to targeted permanent modifications. Recent advancements have significantly improved the efficiency and specificity of CRISPR/Cas9, making it a versatile tool in both research and therapeutic contexts. Recently Doudna and Charpentier adapted the CRISPR/Cas9 system for GE by combining the crRNA and tracrRNA molecules into what is now called a single guide RNA[34] (sgRNA; **Figure 10B**).

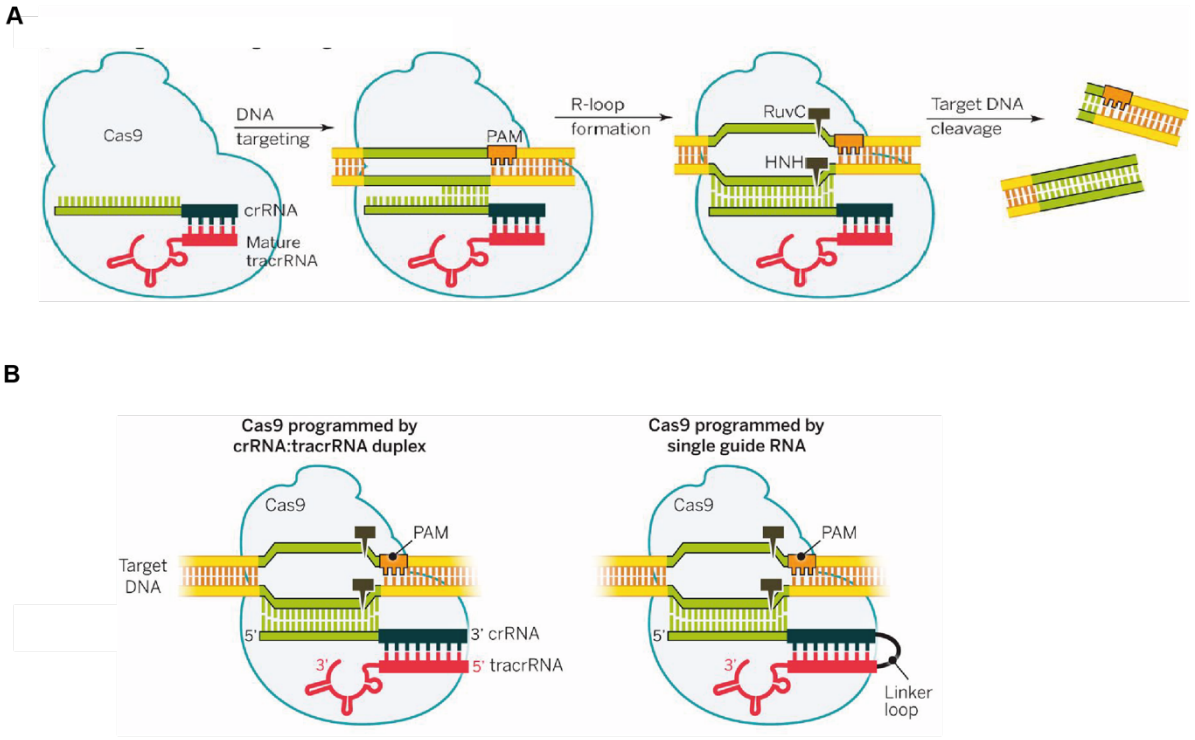


Figure 10. CRISPR/Cas9 recognition of the target sites. A) CRISPR/Cas9 binds to a tracrRNA and a crRNA and recognizes its target site, which is adjacent to a PAM sequence. RuvC and HNH domains cleave target and non-target strands respectively to generate a DSB. **B**) sgRNA design for CRISPR/Cas9 targeting. Adapted from Doudna and Charpentier [34].

In the last years, the development of high-fidelity Cas9 variants, such as SpCas9-HF1 and eSpCas9, has greatly reduced off-target effects, enhancing the safety profile of GE [91,92]. Moreover, base editing technologies, which use a modified Cas9 enzyme that nicks only one strand of DNA, allow for precise nucleotide substitutions without inducing double-strand breaks, offering a gentler and more accurate method for genetic correction [93]. Importantly, CRISPR/Cas9 has emerged as a transformative tool also in the context of AAVs. This technology, alone or in combination with knocking in strategies is being explored extensively in different contexts [65,94–96].

Double-Strand Breaks induced- DNA repair mechanisms.

When a double-strand break (DSB) occurs within the DNA, cells primarily use two distinct repair mechanisms to repair the damage: Homology-Directed Repair (HDR) and Non-Homologous End Joining (NHEJ). Each pathway plays a crucial role in maintaining genomic integrity, yet they differ in their processes and outcomes significanlty. Understanding these pathways provides crucial insights into cellular responses to DNA damage and informs therapeutic strategies for genetic diseases and cancer. HDR is a high-fidelity repair pathway active primarily during the S and G2 phases of the cell cycle which ensures accurate repair by using a homologous sequence to guide the process. To start the proces the MRN complex (MRE11-RAD50-NBS1) initiates end resection at the DSB, creating 3' single-stranded DNA (ssDNA) overhangs [97,98]. This resection is extended by exonucleases like EXO1 and helicases such as BLM [99]. The ssDNA is then coated by the Replication Protein A (RPA), ans subsequently by RAD51 that forms nucleoprotein filaments around the ssDNA, facilitating the search for a homologous sequence on the sister chromatid[100]. The RAD51-ssDNA nucleoprotein filaments invade the homologous double-stranded DNA to form a displacement loop (D-loop). Next the DNA polymerase extends the 3' end of the invading strand using the homologous template to accurately repair the break [101]. Finally during the resolution step, the extended DNA ends are captured, and the D-loop is disassembled. The newly synthesized DNA is then ligated, leading to crossover or non-crossover outcomes, which ensures the precise restoration of the original DNA sequence without introducing mutations, thus preserving genomic integrity [102,103] (**Figure 11**). HDR has been exploited extensively for gene correction, by delivering an exogenous donor sequence with homology arms that can be used as a template for the DSB repair. For ex vivo applications, the most widely used templates are single-stranded oligodeoxynucleotides (ssODNs) with at least 40bp of homology on each side [104]. This approach has been used extensively to modify hematopoietic stem cells as well as in mice [105–107].

Moreover, in the last years HDR has also been used for in vivo gene editing using a donor dsDNA with homology arms flanking the sequence to knock-in [108]. Differently from HDR, NHEJ is the most prevalent DSB repair pathway in mammalian cells active throughout the cell cycle, particularly in the G1 phase. NHEJ is known for its rapid response, with no need of homologous sequence, making it an error-prone process. The first step involeves the recognition of the DBS. Upon a DSB, the Ku heterodimer (Ku70/Ku80) quickly binds to the DNA ends. This binding protects the DNA from degradation and recruits the core NHEJ machinery, including the DNA-dependent protein kinase catalytic subunit (DNA-PKcs) [109]. Next, the DNA-PKcs forms a complex with Ku at the DNA ends, activating the nuclease Artemis, which processes the DNA ends by trimming or adding nucleotides to prepare them for ligation [110]. The following step is the ligation: the DNA Ligase IV-XRCC4 complex ligates the processed DNA ends. Due to the imprecision of end processing, this can lead to mutations or small indels [109,111]. The re-ligated DNA ends restore the continuity of the DNA, albeit often with small errors, which can lead to chromosomal translocations or other genomic rearrangements (**Figure 11**)[109].

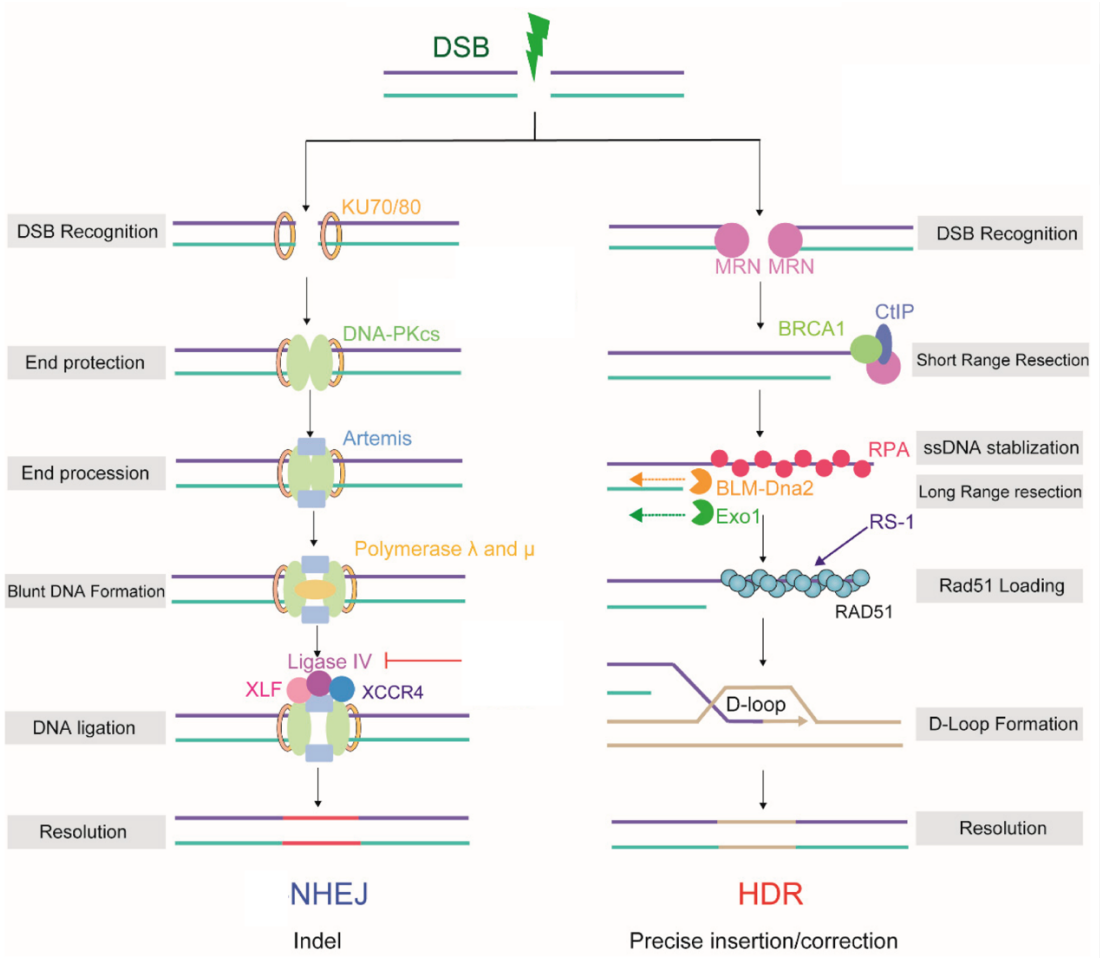


Figure 11. The mechanism and outcome of two major repair pathways: NHEJ and HDR in mammalian cells. Key factors are reported for each repair pathway**.**

Homology-independent targeted integration (HITI)

Different groups attempted to use NHEJ for knock-in by exploiting its propensity to rejoin DNA ends without the need for homologous regions in a donor template, making it a versatile tool for genome editing with different genome editing machinery [112]. However, these approaches faced an important drowback which consisted in the limited ability to control the direction of the integration. Given the lack of flanking homology arms, NHEJ was able to integrate the donor DNA in both directions, lowering the efficiency of correct transgene integration, generating undesired integration products[113]. To overcome this shortcoming, in 2016 Suzuki et al. developed a powerful CRISPR/Cas9-based platform called Homology-independent targeted integration (HITI)[114]. To date HITI has showed promising results in both newborn and adult tissues[95,115]. The power of this approach lies in its design. The HITI donor DNA consists of a promoter-less CDS of a desired transgene to integrate flanked, at its 5' and 3' extremities, by the same gRNA sequence of the endogenous target site but in an inverted orientation. The presence of these inverted gRNA sequences, allow to control the direction of the integration. Upon CRISPR/Cas9-mediated cleavage at both the endogenous locus and at the extremities of the HITI donor DNA, the NHEJ repair pathway promotes donor DNA integration at the nuclease-induced DSBs. Once proper integration has occurred, the flanking gRNA sites are destroyed preventing a further nuclease cleavage. This allows the expression of the therapeutic transgene from the endogenous promoter (**Figure 12**).

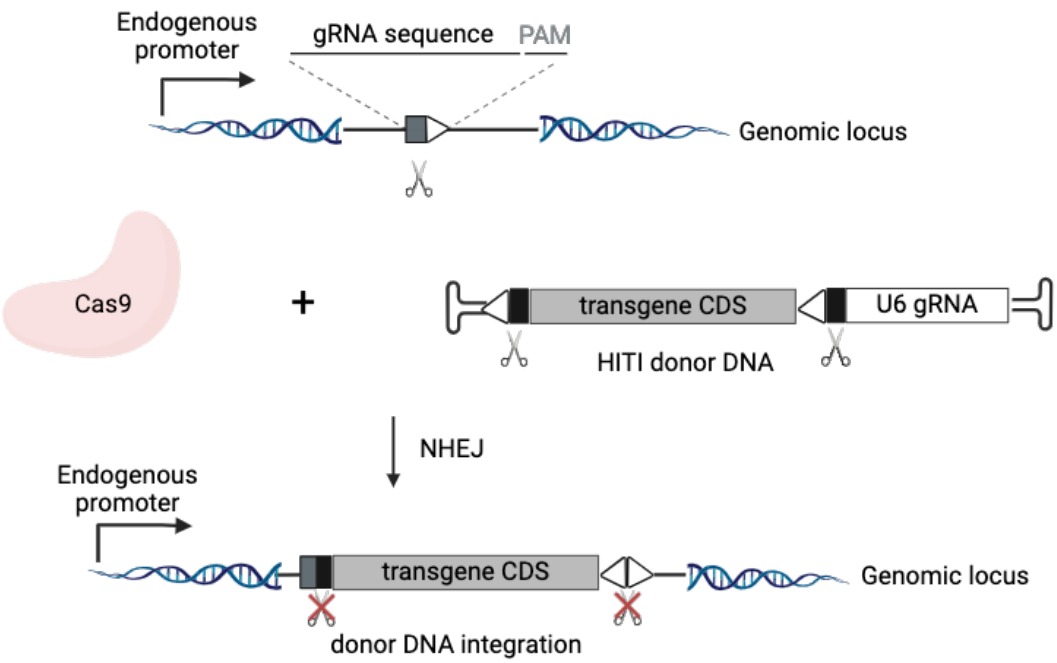


Figure 12. Schematic representation of the homology-independent targeted integration (HITI) platform. The gRNA sequence is designed within the genomic target site. The nuclease SpCas9 is depicted. The AAV contains the HITI donor DNA with the desired promoter-less transgene coding sequence (CDS); U6 gRNA, expression cassette for the gRNA (depicted); gRNA and PAM sequences are represented with a black (inverted gRNA sequences at the extremities of the donor DNA) and grey (genomic locus) boxes and white triangle, respectively. The non-homologous end joining (NHEJ) repair pathway of the cell leads to the integration of the cleaved donor DNA at the on-target site. The integrated transgene will be expressed by the endogenous promoter.

AIMS

The overall objective of my PhD project was to develop innovative AAV-mediated liver-directed gene therapy platforms for the treatment of haemophilia A and mucopolysaccharidosis type VI with two specifics aims:

The first aim was to overcome the limitation of the AAV transfer capacity, by developing a strategy based on protein trans-splicing (PTS) mediated by split-inteins. This platform expands the AAVs cargo capacity allowing for the treatment of those conditions caused by mutations in large genes. PTS involves the use of two separate AAV vectors, each containing regular size and well-defined genomes that work together to reconstitute the full-length therapeutic protein inside the target cells. To evaluate PTS efficacy, I used a mouse model of HemA knock-out for the large Factor 8 (F8). In parallel, **the second aim of my project was focused at expanding the AAV potential to neonatal treatments by overcoming the AAV genome dilution in the highly proliferative neonatal liver tissue**, **to prevent the loss of the therapeutic transgene overtime**. To achieve this, I developed a CRISPR/Cas9-based platform that utilizes homology-independent targeted integration (HITI) of a therapeutic donor DNA into a specific genomic locus, mediated by AAV vectors. HITI hijakcs the non-homologous end joining (NHEJ) able to target both newborn and adult tissues, enabling transgene integration following a single AAV-HITI administration. To evaluate the efficacy of this platform, I employed two mouse models of inherited genetic diseases: a model of HemA and a model of mucopolysaccharidosis type VI (MPS VI), a lysosomal storage disease caused by a deficiency of the secreted, arylsulfatase B (ARSB) enzyme.

The development and validation of these novel AAV-based platforms could significantly expand therapeutic options for those diseases that require improved delivery methods or early intervention.

MATERIALS AND METHODS

Generation of the AAV vector plasmids

The plasmids used for AAV intein vector production derived from the pTigem AAV plasmid that contains the ITRs of AAV serotype 2 [116]. The enanched green fluorescent protein (eGFP) coding sequence was split at Cysteine71 while the F8-N6 protein was split at Serin 962 (Ser962) considering the signal peptide exclusively within the N6 linker (in place of the B domain) to preserve the integrity of the other more critical protein domains. Further split considerations were taken into account based on the intrinsic amino acid residue requirements for efficient protein *trans*-splicing with the *Npu* inteins. In particular, the main prerequisite is the presence of an amino acid containing either a thiol or hydroxyl group (Cysteine, Serine or Threonine) as the first residue in the 3' half of the coding sequence [117,118].

Split-inteins included in the plasmids were the split-inteins of DnaE from *Nostoc punctiforme* (*Npu*) [119]. The plasmids used in the study were under the control of either the ubiquitous cytomegalovirus (CMV) promoter tornabene 2019 or the liver-specific hybrid liver promoter (HLP, [120] or the human thyroxin binding globulin (TBG) promoter [121]. The polyadenylation signal (polyA) used was either the short synthetic polyA [122] or the bovine growth hormone (BGH) polyA signal. For the generation of the heterologous split-inteins, the same splitting point (Ser962) was used. The N-split intein flanking the 5' half plasmid was the N-intein of DnaB from *Rhodothermus marinus* (*Rma*, [81,82] while the C-split intein flanking the 3' half plasmid was the C-intein of the DnaE. Heterologous split-intein plasmids were produced under the control of the ubiquitous cytomegalovirus (CMV) promoter since they were only used for *in vitro* purposes. The codon optimised BDD-F8-V3 plasmid used as golden standard, was kindly provided by Dr Amit C. Nathwani [120]. For the HITI plasmids the mouse albumin (m*Alb*) gRNA (5'-GTATTTAATAGGCAGCAGTG-3') was

selected using the benchling gRNA design tool (www.benchling.com), targeting the intron 13 of the albumin locus considering the best predicted on-target and off-target scores. The scramble RNA was designed so as not to align with any sequences in the mouse genome.

AAV vector production and characterization

AAV serotype 8 vectors (AAV8) were produced by InnovaVector s.r.l by triple transfection of HEK293 cells followed by two rounds of $CsCl_2$ purification [123]. For each viral preparation, vector titres (genome copies/ml) were determined by averaging the titre achieved by dot-blot and by TaqMan-PCR (Cat#4304437, Applied Biosystems, Carlsbad, California, USA) quantification analysis.

Transfection of HEK293 cells

HEK293 cells were maintained and transfected using the calcium phosphate method (1 μg of each plasmid/well in 6-well plate format) as already described [124].The total amount of DNA transfected in each well was kept equal by addition of a scramble plasmid when necessary. 12 hour post-transfection (hpt) medium was switched to Opti-MEM reduced serum medium (31985062; Gibco, ThermoFisher Scientific, Germany) 1ml/well until the time point of 72 hpt when cells were harvested and medium was collected.

Western blot analysis

Samples (HEK293 cells or liver lysates) were lysed in RIPA buffer to extract F8 protein. Lysis buffer was supplemented with protease inhibitors (11697498001; Roche, Basel, Switzerland) and 1mM phenylmethylsulfonyl. For medium samples, upon cell harvesting at 72 hours post-transfection (hpt) medium samples were centrifugated at 4°C for 15 minutes to remove cell debris. Purified medium was collected; 30 µl of medium mixed with 1X Laemmli sample buffer. All samples were denatured at 99°C for 5 minutes in 1X Laemmli sample buffer. Liver lysates (100 µg), cell lysates (50 µg) and medium samples were separated by either 12% (for excised intein detection) or 6% (for full-length F8 protein and single halves detection) SDS–polyacrylamide gel electrophoresis (SDS-PAGE). The antibodies used for immuno-blotting are as follows: anti-3xFlag (1:2000, A8592; Sigma-Aldrich, Saint Louis, MO, USA) to detect the full-length F8-N6 protein and both the 5' and the 3' halves; anti-β-Actin (1:1000, NB600-501; Novus Biological LLC, Littleton, CO, USA) to detect β-Actin proteins which were used as loading controls for the 12% SDS-PAGE; anti-Calnexin (1:2000, ADI-SPA-860; Enzo Life Sciences Inc, New York, NY, USA) to detect Calnexin, used as loading controls for the 6% SDS-PAGE. The quantification of the full-length F8-N6 bands detected by Western blot was performed using ImageJ software (free download is available at http://rsbweb.nih.gov/ij/).

Protein concentration in AAV-HITI-treated sera samples collected from HemA mice was determined by Bradford Reagent (Cat#5000006, Bio-Rad). Ten micrograms of sera were denatured at 100°C for 5 min in 1× Laemmli sample buffer supplemented with 1 M dithiothreitol (DTT). Next, protein samples were resolved on a 4–15% SDS–PAGE and transferred onto polyvinylidene difluoride (PVDF) membrane. After transfer to PVDF membrane, blot was blocked with TBS-Tween-20 containing 5% non-fat milk for 1 h at room temperature followed by incubation with primary antibody (Cat#NBP259627H,Novus Biologicals;

dilution: 1/2,000) 1h at 4°C. Ponceau (Cat#P7170-1L, Merck) was used as normalizer in sera blot. Liver specimens were mechanically homogenized using metal beads and lysed in RIPA buffer, supplemented with protease inhibitor cocktail (Cat#78430, Merck). Samples were incubated for 30 min on ice, vortexed every 10 min, and centrifuged at 16,200 *g* for 20 min. Pellets were discarded and lysates were used for Western blot analyses. After lysis, protein samples were treated as mentioned above. Hundred micrograms of liver proteins were loaded for each specimen into a 6% SDS–PAGE; after transfer to PVDF membrane, blot was blocked with TBS-Tween-20 containing 5% non-fat milk for 1 h at room temperature followed by incubation with primary antibody overnight at 4°C. The primary antibodies used for immuno-blotting were: rabbit anti-*Sp*Cas9 (Cat# MA1-201, Thermo Fisher Scientific; dilution: 1/1,000) mouse anti-Calnexin (Cat#ADI-SPA-860F, Enzo Life Sciences; dilution: 1/2,000).

Southern Blot analysis

DNA was extracted from 6×10^{10} viral particles measured as genome copies (GC). To digest unpackaged genomes, the vector solution was incubated with 30 µl of DNase I (04536282001; Roche,Italy) in a total volume of 300 µl, containing 50 mM Tris pH 7.5, and 1 mM $MgCl_2$ for 2 hours at 37°C. The DNase was then inactivated with 50 mM EDTA, followed by incubation at 50°C for 1 hour with proteinase K and 2.5% *N*-lauryl-sarcosil solution to lyse the capsids. The DNA was extracted twice with phenol-chloroform and precipitated with 2 volumes of ethanol 100% and 10% sodium acetate (3 M) and 1 µl of Glycogen (1090139300; Roche,Italy) was performed as previously described[125]. Single-stranded DNA was quantified with Qubit® ssDNA Kit (Q10212; ThermoFisher Scientific, Germany). A probe specific for the HLP promoter was used;1.4 $x10^{10}$ GC for the single vector AAV-Codop V3 and for both the 5' and the 3' AAV-N6 intein were loaded on an alkaline agarose gel electrophoresis.

Immunoprecipitation and Mass Spectrometry analysis

Cells were plated in 100 mm plates (5x105 cells/plates) and transfected with either the single Codop N6 plasmid or Codop N6 intein plasmids using the calcium phosphate method (20 μg of each plasmid/plate). Cells were harvested 72 hpt and both the single CodopN6 and the CodopN6 intein proteins were immunoprecipitated using anti-flag M2 magnetic beads (M8823; Sigma-Aldrich), according to the manufacturer instructions. Proteins were eluted from the beads by incubation for 15 minutes in sample buffer supplemented with 4 M urea at 37°C and 10minutes at 99°C. Samples were then loaded on a gradient 4-10% SDS–polyacrylamide gel electrophoresis. In total 8 protein bands (from HEK293 cells transfected 5 times independently with CodopN6 intein plasmids) were cut after staining with Instant Blue (ISB1L; Sigma-Aldrich) and were used for protein sequencing. Briefly, 8 gel slides were used for digestion by the following enzymes: Lysin and Trypsin. The resulting peptides were identified using nanoscale Liquid Chromatography coupled to tandem Mass Spectrometry (nano LC-MS/MS) analysis. Data obtained were processed using MaxQuant and the implemented Andromeda search engine.

Systemic vector administration

Retro-orbital injection in the F8 intein studies were approved from the Italian Ministry of Health; department of Public Health, Animal Health, Nutrition and Food Safety number 379/2019-PR. Adult C57Bl/6 mice (5 weeks of age) were retro-orbitally injected with AAV8 at the dose of 5 x 10^{11} GC of each vector per animal. Adult knockout males (between 7 to 11 weeks of age) were retro-orbitally injected with AAV8 with either AAV-N6 intein or the single AAV-CodopV3 as a positive control at the dose of 5 x 10^{11} GC of each vector per animal. AAV-CodopN6 intein were retro-orbitally injected with the dose of 1.5 x 10^{11}GC of each vector per mouse.

AAV-HITI animals studies were carried out in accordance with the Italian Ministry of Health regulation for animal procedures (Ministry of Health authorization number: 352/2020-PR and 626/2022-PR). The injections were performed under general anaesthesia, as previously described [115]. Temporal vein injections in neonatal mice (C57BL/6J or HemA) were performed at post-natal p1-2 following the protocol published by Gombash Lampe et *al*.[126]. For adult treatments, (C57BL/6J, or HemA) retro-orbital injections were performed. The following doses were used in the experiments aimed at evaluating HITI efficiency in C57BL/6J newborn mice: high doses 1.2 × 10^14 total genome copies (GC/Kg; 6 × 10^13 GC/Kg for each vector); medium doses 3.9 × 10^13 total GC/Kg (1.95 × 10^13 GC/Kg for each vector); low doses 1.2 × 10^13 total GC/Kg (6 × 10^12 GC/Kg for each vector). For rescue experiments in neonatal HemA mice, we used 3.9 × 10^13 total GC/Kg (1.95 × 10^13 GC/Kg for each vector). For rescue experiments in HemA adult mice, we used 1.2 × 10^13 total GC/Kg (6 × 10^12 GC/Kg for each vector). AAV8 vectors were used for all *in vivo* studies because of their high liver transduction efficiency [127].

Plasma collection and F8 assays

Briefly, nine parts of blood were collected by retro-orbital withdrawal into one part of buffered trisodium citrate 0.109M (5T31.363048; BD, Franklin Lakes, NJ, USA). Blood plasma was collected after samples centrifugation at 3000 rpm at 4°C for 15 minutes.

To evaluate F8 activity chromogenic assay was performed on plasma samples using a Coatest® SP4 FVIII-kit (K824094; Chromogenix, Werfen, Milan, Italy) according to manufacturer's instructions. Standard curve was generated by serial dilution of commercial human F8 (Refacto, Pfizer). Results are expressed as International Units (IU) per deciliter (dl).

Activated partial thromboplastin time (aPTT), was measured on plasma samples with Coatron M4 (Teco, Bünde, Germany) using the aPTT program following the manufacturer's manual.

To quantify F8 antigen levels the ELISA kit (FVIII-AG;VisuLize FVIII ELISA kit, Affinity Biologicals,Arcore,Italy) was used according to manufacturer's instructions.

Tail-clip assay

Mice were anesthetized and the distal part of the tails was cut at 2-3 mm of diameter and immediately put in a prewarmed 0,9% saline solution and allowed to bleed for 10 minutes without disturbance. Tails were then cauterized, and mice were sacrificed. The mixture of collected blood and physiological saline solution was centrifuged at 1,500g for 5 minutes. The collected erythrocytes were lysed with water and the hemoglobin content was measured at an optical density of 416 nm.

Indirect enzyme-linked immunosorbent assay (ELISA) to detect anti-F8 antibodies

To evaluate the presence of anti-F8 antibodies an indirect enzyme-linked immunosorbent assay (ELISA) was performed using ZYMUTEST^{TM}Anti-VIII Monostrip IgG (RK039A; HYPHEN BioMed, France) according to manufacturer's instructions. The secondary antibody used to detect mouse IgG is the goat-anti-mouse IgG (H+L) HRP conjugate (1:3000, AP308P; Sigma-Aldrich).

Bethesda assay to detect F8 inhibitors

Plasma samples containing high levels of F8 were heat inactivated at 58°C for 90 minutes. A standard curve was generated by serial dilution of the commercial human F8 (Refacto, Pfizer) in F8 deficient plasma. Serial dilutions of all samples were done in F8 deficient plasma and then mixed 1:1 with 100% Refacto (Pfizer) diluted in F8 deficient plasma. All experimental samples and controls were then incubated at 37°C for 2 hours. After that, all samples and the standard curve were analyzed by aPTT following the manufacturer's manual.

Cytokines and chemokines assay

Serum samples were analyzed with mouse cytokine & chemokine 36-Plex ProcartaPlex 1A Panel (EPX360-26092-901;Thermofisher) according to manufacturer's instructions.

Liver fluorescence imaging

To evaluate Ds-Red expression in liver, C57BL/6J mice were injected at post-natal day (p) 1-2. Livers were harvested at p30 under anaesthesia as previously described [95,115]. A small piece of each lobe was dissected, fixed in 4% PFA overnight, infiltrated with 15% sucrose over the course of a day and 30% sucrose overnight before being included in O.C.T. matrix (Cat# 0782, Kaltek) for cryo-sectioning. Five-µm-thick liver cryo-sections were cut, distributed on slides, and mounted with Vectashield with DAPI (Cat#H-1200-10, Vector Lab). Cryo-

sections were analyzed under a confocal LSM-700 microscope (Carl Zeiss), using appropriate excitation and detection settings for Ds-Red and DAPI. For assessment of HITI efficiency in mouse liver cryo-sections, three to four images of each liver were acquired at 20x magnification and then analyzed using ImageJ (Fiji) software (http://rsbweb.nih.gov/ij/) as previously described[128]. We counted a minimum of 900 hepatocytes, identified by DAPI staining of the nucleus, for each image. The hepatocytes expressing the Ds-Red as result of the integration were unequivocally identified and counted based on their shape. The final value (~15%) was then obtained by dividing the number of Ds-Red positive hepatocytes over the total number of DAPI cells and multiplied by 100 to obtain the percentage (%).

In situ hybridization

Liver cryo-sections were used for the in situ hybridization (ISH) with the technology of Basescope™ (Advanced Cell Diagnostic, USA) in accordance with the manufacturer's protocol. ISH labelling of HITI donor DNA was performed using a 3-ZZ paired probe (sense) (Cat# 1134401, ACD). ISH was performed using the BaseScope Duplex Reagent Kit (Cat# 323871, ACD). Custom pre-treatment conditions included target retrieval of 30 min at 95 C–100 C, RNAscope Protease III (Cat # 322380, ACD) for 30 min at room temperature (RT). Custom counterstaining with Mayer haematoxylin for 1min at RT was performed after probe, amplification, and chromogen steps in the kit assay. Sections were analyzed with scanned with ZEISS Axio Scan.Z1. The whole digital slides were viewed by zen blue software. Red positive spots of HITI DNA were quantified with QuPath software. For the quantitative analysis of positive signals, we selected 4-6 liver regions at 20x magnification from mice belonging to AAV-HITIgRNA, -scRNA and PBS experimental groups, and the results were expressed as a percentage of positive nuclei over the total nuclei analyzed.

Serum albumin measurement

Blood was collected at p360 from AAV-HITI-treated and control mice via eye bleeding and centrifuged as previously described[128]. Serum samples were diluted 1:30.000 and analyzed with a mouse albumin ELISA kit (Cat#108791, Abcam, Cambridge, UK) following the manufacturer's instructions.

Histopathological analysis

Right, left, median, and caudal liver lobes were embedded in paraffin and sectioned at a thickness of 5 μm. To assess histological features, Haematoxylin/Eosin (Diapath) staining was performed according to standard protocols and samples were analyzed by an expert histopathologist from Histopathology Unit (Cogentech Ltd. Benefit Corporation, Milan) in blind.

DNA extraction

DNA extraction was performed using the DNeasy Blood & Tissue kit (Cat# 69504, QIAGEN) following the manufacturer's instructions.

Surveyor Assay

1-3uL of the PCR products (according to PCR efficiency) was used for Surveyor Assay, following the GeneArt™ Genomic Cleavage Detection Kit manufacturer's recommendations. In short, DNA was de-annealed at 99C and re-annealed by a slow temperature gradient in a thermocycler. After re-annealing, 1uL of Detection Enzyme (T7Endonuclease) was added, and samples were incubated for an hour at 37 degrees. After incubation, samples were run in a 2% agarose gel in order to detect DNA cleavage products resulting from INDEL presence.

Tracking of INDELs by Decomposition

PCR products were also used for Sanger sequencing. Sequences were then used for TIDE software (https://tide.deskgen.com/) and ICE analysis (https://ice.synthego.com/) of INDEL frequency.

Quantitative PCR (qPCR)

Viral genome copies (GC) were measured by qPCR analysis in 100ng of genomic DNA extracted from AAV-HITI-treated liver samples at 1-week, -month and 1-year upon AAV-HITI administration, using the LightCycler 480 SYBR Green I Master mix (Cat#04707516001, Roche). The following primers: forward HLP: 5'-CTCCTCCGATAACTGGGGTGAC-3' and reverse HLP: 5'-GCCCTGTCCTCGTCCGTATTTA-3', were used to detect AAV8 -*Sp*Cas9 vector; forward BGH 5'-TCTAGTTGCCAGCCATCTGTTGT-3'; reverse BGH 5'-TGGGAGTGGCACCTTCCA-3' were used to detect the AAV-HITI donor DNA following the qPCR protocol from Roche. Briefly, enzyme activation was set at 95 °C during 20 sec -3 min (1 cycle), followed by the denaturation step at 95 °C during 3 sec and annealing extension and acquisition at 60 ºC at 20 sec with the last two points set at 40 cycles.

CAST-Seq

To identify chromosomal aberrations, nominate off-target sites and analyze integration events at the on-target site, we performed CAST-Seq[129]. High-throughput sequencing libraries were essentially prepared according to Turchiano et al., 2021[129]. In search for off-targets, a third decoy primer was included to prevent predominant amplification of the target locus after successful HITI. In total, three CAST-Seq libraries from liver genomic DNA samples extracted from three different 1 year-old MPS VI mice treated as newborn with high doses of AAV-HITI were prepared and compared to a single library generated from control mice samples (scRNA). In order to analyze in depth the integration of AAVs and parts thereof, two CAST-Seq libraries of HITI-treated and a single library of a

control mouse were sequenced. For these samples, the HITI-specific decoy primer was not used during library preparation. Sequencing of the samples was outsourced to GeneWiz (division of Azenta Life Sciences) who collected 2x150bp paired-end reads using an Illumina NovaSeq 6000 instrument. The bioinformatics analysis was performed using the previously published pipelines [129].

Nanopore long-read sequencing

Fragments for long-read sequencing were amplified from DNA extracted from 1 year-old MPS VI mice treated as newborn with high doses of AAV-HITI using KAPA HiFi Hotstart Polymerase (Cat#07958927001, Roche,). The cycling conditions were: 3 min at 98°C followed by 35 cycles of 98°C for 20 sec, 61°C (5' junction) or 63°C (3' junction) for 30 sec, and 4 min at 72°C. A final elongation of 10 min at 72°C was programmed after the last cycle. Multiple PCR reactions were pooled and the pool subjected to bead purification using 0,9X Ampure XP beads (Cat# A63881, Beckman Coulter) and fragments eluted in a small volume of water (30 µl). The thus concentrated PCR pool was subjected to agarose gel electrophoresis and the prominent PCR product was gel-extracted using the QIAquick® Gel Extraction Kit (Cat# 28706, Qiagen). With the PCR step, 24-nt long barcodes were introduced, in accordance with previously published data [130]. The 5' and 3' junction PCR products were in a next step prepared for sequencing using the Ligation Sequencing Kit V14 (Cat# SQK-LSK114, Oxford Nanopore Technologies) following the manufacturer's instructions. Samples were immediately sequenced on an R10.4.1 MinION Flow Cell (Cat#FLO-MIN114, Oxford Nanopore Technologies) using a MinION sequencing device (Cat#MIN-101B, Oxford Nanopore Technologies) operated with MinKNOW software version 23.04.6. Reads were demultiplexed by employing a custom pipeline that allows for up to 4 mismatches in the barcode sequence using pcregrep and GNU Parallel [131]. We decided to allow up to 4 mismatches based on previously

published data [132] as a compromise between a low false annotation rate versus loss of valuable information. Of note, the two samples described here were analyzed together during one sequencing run in the absence of other samples thus reducing the chance of wrongful read assignment. Obtained reads were aligned using Minimap2 v. 2.24-r1122 [133] and further processed with Samtools [134]. ITR sequence insertions were counted from the CIGAR string in a window of 200 bp around the integration site junctions using a custom script.

Short-read sequencing

Assessment of Indels in short sequencing reads was performed using CRISPResso2 [135]. For the quantification of ITR sequence integration, raw FASTQ files were processed using BBduk (https://sourceforge.net/projects/bbmap/) and aligned to the reference amplicon using BWA MEM v0.7.17 [136]. Correctly mapped reads were extracted from the FASTQ file and aligned to the ITR sequence. The total number of all alignments and the number of all alignments to the ITR sequence were quantified and the relative frequency was calculated.

Off-target site analysis

The top 10 predicted off-target sites were identified using the CRISPOR web tool (CRISPOR.org) based on GRCm39/mm39 mouse genome reference, and sorted by CFD off-target scores. Equal amounts of genomic DNA extracted from 3 AAV-HITI-gRNA-treated livers and 3-scRNA-treated livers were used to amplify between 150 to 300 bp genomic regions flanking the off-targets. PCR products were amplified with specific primers and quantified using Qubit 4.0 fluorometric Assay (Thermo Fisher Scientific). An equal amount of each PCR product from each liver sample was pooled together. Next-generation sequencing (NGS) amplicon library were preprepared from 10 ng of pooled PCR products using the NEGEDIA DNAseq Low Input sequencing service (NEGEDIA s.r.l) which included library preparation, quality assessment and sequencing on a NovaSeq 6000 sequencing system using a paired-end, 2x150 cycle strategy (Illumina Inc.).

The resulting FASTQ files were then analyzed using CRISPRessoV2 [135] using the off-targets sequences as reference for analysis. A 17-nucleotide window (upstream and downstream of the cleavage site) was considered for evaluation. NGS analysis was performed by NEGEDIA s.r.l.

Serum ARSB enzymatic activity

Blood samples were collected at different timepoints from AAV-HITI-treated and control MPS VI mice as previously reported. Serum ARSB activity was measured by an immune capture assay based on the use of a specific custom-made anti-hARSB polyclonal antibody (Covalab, Villeurbanne, France) following a previously described protocol [25].

GAG level analysis in urine and tissues

Urine samples were collected over 24 h using metabolic cages at p60, p90, p180, p270 and p360 from MPS VI-treated and control mice. Samples were centrifuged briefly to remove debris and diluted 1:50 in water to measure GAGs content. Fifty µl of diluted urine or 250 µg of protein lysate were then used for GAGs evaluation as previously reported [95].

Alcian blue staining in histological sections

After methacarn fixation, livers, kidneys, spleens, and hearts were dehydrated by immersion in increasing concentrations of alcohol (70%, 80%, 90%, 100%) and then in Sub-X. All tissues were embedded in paraffin and sectioned transversally into 7-µm-thick serial sections on a microtome. Tissue sections were de-paraffinized, rehydrated, then washed in water and stained with 1% Alcian blue (Cat#A5268-25G, Merck) in hydrochloric acid (5 minutes for hearts, 60 minutes all remaining tissues). Counter-staining was performed with 0.1% Fast-Red (Cat#N8002-5G, Merck) in water for 1 minute. Liver, kidney, and spleen sections stained with Alcian Blue were imaged with Leica Microscope DM5500. Heart sections stained with Alcian Blue were scanned with ZEISS Axio Scan.Z1. The whole digital slides were viewed by Zen Blue software

(https://www.zeiss.com/microscopy/en/products/software/zeiss-zen-lite.html). Quantitative analyses of Alcian Blue staining in myocardial tissue and in mitral valves were performed by Qpath software (https://qupath.github.io/). Alcian Blue quantification in myocardium was measured in two fields of identical area. Alcian Blue quantification in the mitral valve was performed on the entirety of the valve. Results are expressed as Alcian blue positive area/total area.

Bone analysis

Radiography images were performed on AAV-HITI-treated MPS VI mice and unaffected controls were anesthetized with an intraperitoneal injection of ketamine (10 mg/Kg) combined with medetomidine (1 mg/Kg). Skull weight and length, and tibia and femur length, were measured using ImageJ (Fiji) software (http://rsbweb.nih.gov/ij/).

Statistical analysis

For all the statistical analysis, we performed the Shapiro-Wilk test to check that each condition had followed the normal distribution (null hypothesis). To assess significant differences between two conditions the non-parametric Mann-Whitney test was used in the case of rejection of the null hypothesis (p-value < 0.05). If the null hypothesis was not rejected (p-value ≥ 0.05) the parametric unpaired T-test was applied and the F-test was used to check the homoscedasticity between the two compared conditions. In case of rejection of the null hypothesis (p-value < 0.05) the parametric Welch's t-test was applied. For multiple comparisons and post hoc tests, non-parametric Kruskal-Wallis test was used in the case of rejection of the null hypothesis (p-value < 0.05). The parametric one-way ANOVA test was applied in case of not rejection of the null hypothesis (p-value ≥ 0.05). In this case, we also performed the Brown-Forsythe test to check the homoscedasticity between the compared conditions (null hypothesis): we used the parametric Welch's one-way ANOVA in case of rejection of the null hypothesis (p-value < 0.05). For completeness, we computed the p-values with

post hoc tests for the pairwise multiple comparisons: Tukey's test for one-way ANOVA, Dunnet's test for Welch's one-way ANOVA, and Dunn's test for Kruskal-Wallis test. P-values are represented as follows: * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001; **** p-value < 0.0001.

RESULTS

Specific aim 1: The first objective of my project has been focused at overcoming the limited AAV cargo capacity. To this aim I took advantage of the protein *trans*-splicing mechanism which involves two separate AAV vectors with regular size and well-defined genomes.

Comparison between single AAV and the AAV intein-EGFP system in the liver of wild-type mice

Tornabene et al. [81] previously showed the efficiency of the AAV-intein mediated protein trans-splicing (PTS) in the retina of different models (i.e. mouse and pig). To evaluate the efficiency of the AAV-intein mediated PTS system in the liver, I first adapted the full-length fluorescent eGFP reporter CDS with the DnaE split-inteins from *Nostoc punctiforme* [119]. To optimize the chosen splitting position, I considered the intrinsic amino acid residue requirements for efficient protein trans-splicing with the Npu inteins. Specifically, the main prerequisite is the presence of an amino acid containing either a thiol or hydroxyl group (Cysteine, Serine, or Threonine) as the first residue in the 3' half of the CDS. Therefore, I split the full-length eGFP CDS at Cysteine 71. The first eGFP half (5' CDS) was then fused to the N-terminal half of DnaE at its 3' extremity, while the C-terminal of DnaE was fused to the beginning of the 3' half of the eGFP CDS. Split eGFP constructs were separately cloned under the human thyroxine-binding globulin (TBG) liver promoter [121] and independently packaged into AAV8 vectors that efficiently target the liver. The AAV8-eGFP intein system was delivered by systemic administration in adult C57/Bl6 wild-type mice. A single AAV-carrying the full-length eGFP CDS that fits AAV cargo capacity, with the same regulatory elements present in the intein vectors, was produced for comparison. Mice were injected around 5 weeks of age at the dose of 5×10^{11} GC of each vector per animal. I harvested the livers 4 weeks post-injection (4wpi) and I performed Western blot (WB) analysis of liver lysates. WB quantification of the eGFP bands

showed that AAV-intein reconstitute about 76 % of the full-length eGFP protein produced by the single vector (**Figure 13**).

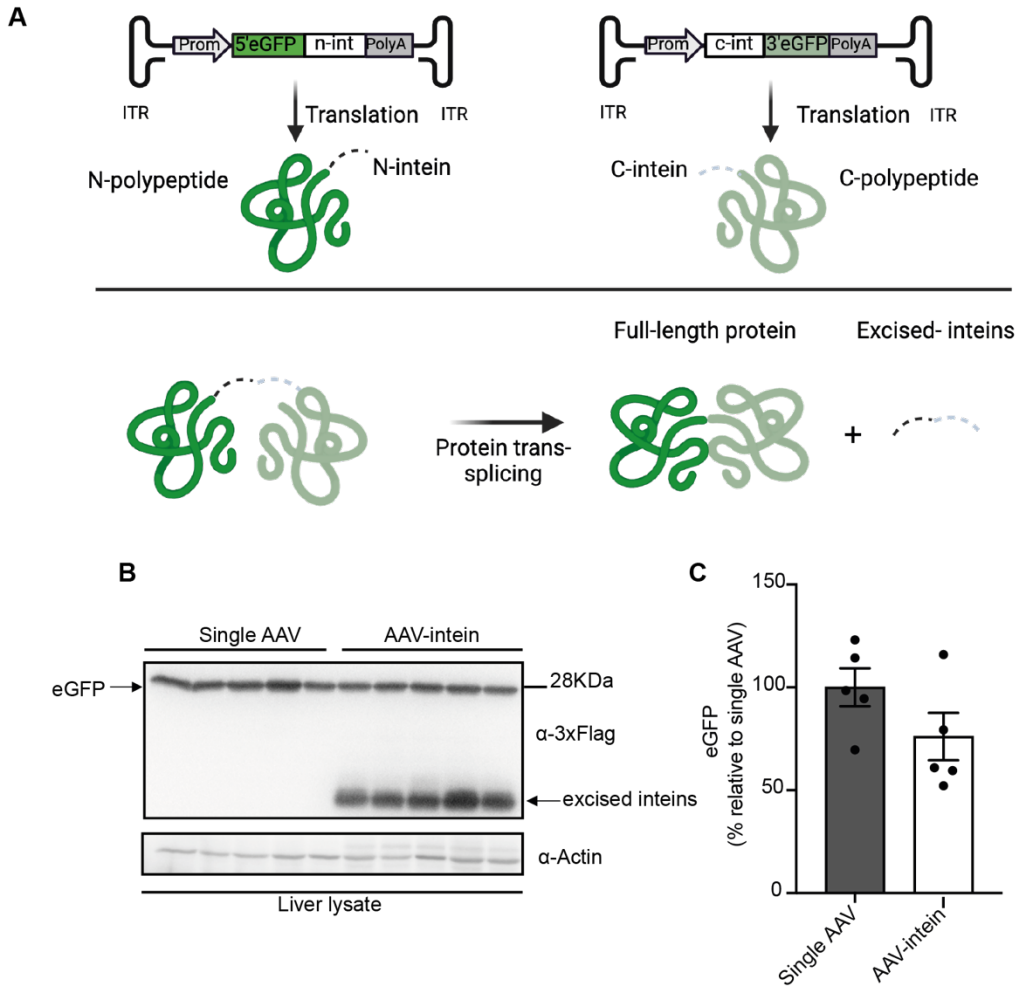


Figure 13. AAV-Intein-mediated eGFP protein trans-splicing in liver. A) Schematic representation of the enhanced green fluorescent protein (eGFP) intein constructs and of the intein-mediated protein trans-splicing. ITR, inverted terminal repeats; Prom, promoter; 5'eGFP,5'eGFP coding sequence (CDS); n-intein, N-terminal of DnaE intein (star symbol), 3xFlag tag; PolyA, short synthetic polyadenylation signal; c-intein, C terminal of DnaE intein; 3'eGFP, 3' eGFP CDS. **B)** Western blot analysis of liver lysates (100 µg) shows that intein-mediated protein trans-splicing efficiently reconstitutes full-length eGFP. Single AAV: n=5; AAV-intein: n=5. Arrows indicate both full-length eGFP and excised inteins. **C**) Quantification of eGFP protein bands.

Characterization of 4 different human F8 variants in HEK293 cells.

Next, I evaluated the activity of 4 different F8 variant in order to select the best to be adapted with the AAV-intein vectors. To this end, I compared the wild-type F8 variant to 3 commonly used B-domain-deleted (BDD, which lack F8 amino acids from 740 to 1649 [137] F8 variants (**Figure 14A**). Specifically, the 3 BDD constructs carry different codon-optimised linkers in the place of the B domain which are designed to promote efficient F8 secretion by mimicking some of the natural F8 post-translational modifications [137]: F8-SQ (SQ) contains the original SQ linker described by Sandberg et al.[138]; F8-N6 (N6) which contains 11 amino acids from the modified SQ activation peptide (SQm) from Ward et al. [139], followed by the N6 human B domain spacer involving 6 (N)-linked glycosylation sites [137]; The F8-V3 variant (V3) which consists of a small 17-aa peptide, that contains the original 6 (N)-linked glycosylation triplets from the N6 inserted into and flanked by the SQ linker [120]. To assess their efficiency, I cloned all four variants independently into the same AAV backbone plasmid under the control of i) the CMV promoter and both a ii) short synthetic polyadenylation signal [122] and iii) a triple flag tag (3xFlag) to allow for protein detection. I then transfected each plasmid in HEK293 cells by transient transfection. Transfected cells were cultured for 12 hours in normal media and then they were kept in serum-free medium until the timepoint of 72 hpt. Cell lysates were used to perform WB analysis and evaluate protein expression (**Figure 14B)**. In parallel, I performed chromogenic assay on the media of the transfected cells to assess the biological activity of each variant. Interestingly, I found that N6 is the most secreted F8 variant (**Figure 14C**). Wild-type F8 had fairly low mean levels of activity of 8.4 International Unit / deciliter (IU/dl), followed by SQ with 27.4 IU/dl, V3 with 57.6 IU/dl, and N6 with the highest mean levels of 71.6 IU/dl. Based on this, I selected N6 as the variant to be tested for further analysis. I therefore adapted this F8N6 variant with AAV-intein. Next, I compared the efficiency of the AAV-N6 intein system with that of a single AAV vector carrying a BDD F8 variant (V3),

which is currently being investigated in clinical trials for gene replacement therapy (NCT03001830)[140,142].

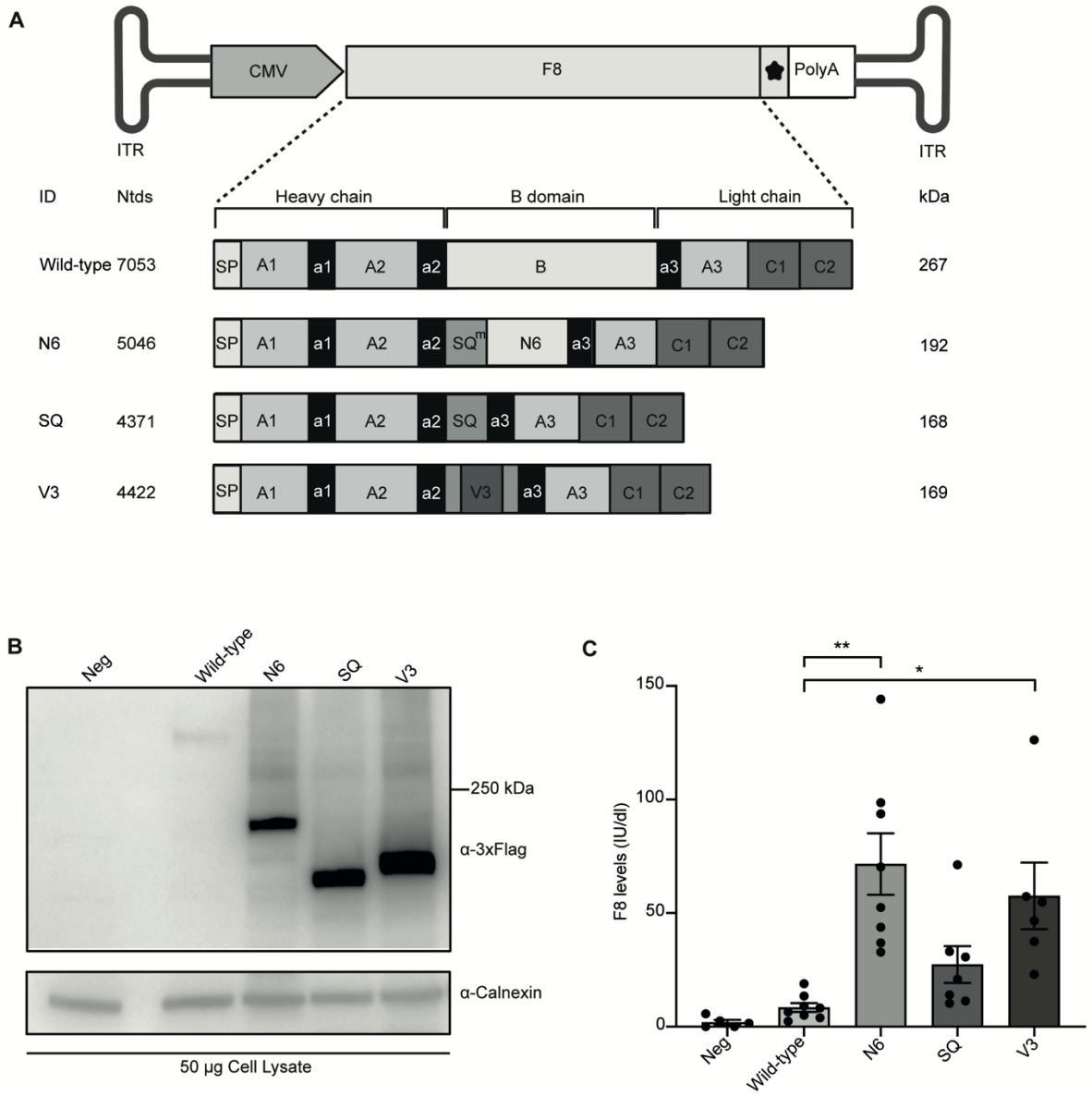


Figure 14. Comparison of human F8 variants in vitro. A) Schematic representation of the four different F8 variants that were cloned into an AAV plasmid: wild-type F8; N6 containing 11 amino acids of the modified SQ amino acid linker (SQ^m) followed by the human N6 B domain; SQ containing the SQ amino acid linker; V3 containing the V3 peptide in the middle of the SQ linker. ITR, inverted terminal repeats; CMV, cytomegalovirus promoter; star symbol, 3xFlag tag; PolyA, short synthetic polyadenylation signal; Ntds, nucleotides; SP, signal peptide. **B)** Western blot analysis of lysates of HEK293 cells 72 hours post-transfection with the various F8 variants. Neg, non-transfected cells. **C)** Chromogenic assay of F8 activity in the

medium of transfected cells are reported as International Units/deciliter (IU/dl). Data are presented as mean±SEM. Significant differences between groups were assessed using The Kruskal-Wallis test followed by the posthoc analysis: Nemenyi's All-Pairs Rank Comparison Test. The Kruskal-Wallis test p value = 2.88e-05.

Characterization of N6 intein sets *in vitro*

First, I split the large F8-N6 coding sequence (CDS) into two fragments i.e. the 5' and 3' half, fused respectively to the N- and C-terminal halves of the DnaE split-inteins from *Nostoc punctiforme*. I designed two different N6 intein sets where the splitting point was selected within the N6 linker (**Figure 14A**) of the N6 variant: in a first set, the full-length N6 CDS was split at Ser962 while in a second inteins set, full-length N6 CDS was split at Ser883, considering the signal peptide. Then I tested the efficiency of each intein system in HEK293 cells. Seventy-two hours post-transfection I harvested the cells and performed WB analysis to assess protein expression. I observed full-length protein expression in the cell lysate of transfected cells with the Ser 962 set (**Figure 15**). Interestingly, the chromogenic assay revealed higher F8 activity levels with the Ser962 N6 intein set than with the Ser883 one. Therefore, I chose the Ser962 N6 intein set for further analysis.

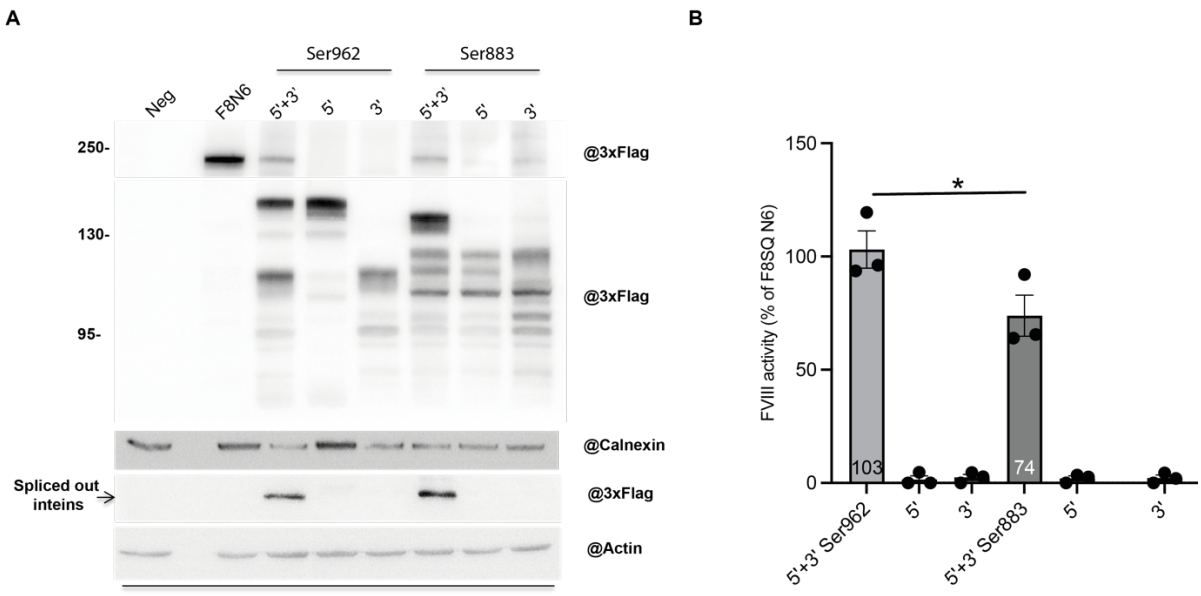


Figure 15. Evaluation of N6 intein sets in HEK293 cells. A) Western blot analysis of 50 ug of cell lysates of HEK293 cells 72 hours post-transfection with the various N6 intein sets. Neg, non-transfected cells. Expected full-length N6 molecular weight is around 190 Kd. The expected molecular weight for the different split-intein sets is as follow: the Ser962_5'half (5') is around 123Kd while the Ser962_3' half (3') is 93Kd.The Ser883_5'half (5') is around 115Kd while the Ser883_3' half (3') is around 100Kd. **B**) Chromogenic assay performed in the medium of transfected cells with different intein sets. Data are reported as International Units/deciliter (IU/dl) and presented as mean±SEM.

To determine if F8 activity in the media of transfected cells was specifically due to the reconstitution of full-length N6 after PTS, I designed a set of N6 constructs flanked by heterologous split inteins. Heterologous split inteins are expected not to self-recognize, thereby preventing the formation of full-length protein. I fused the N-terminal of the 5' half of the N6 CDS to the split N-intein DnaB from *Rhodothermus marinus* (Rma) [81,141], while the C-terminal of the 3' half was fused to the split C-intein DnaE. These intein sets were tested by transient transfection into HEK293 cells. Seventy-two hours post-transfection, I harvested cell lysates and medium and evaluated N6 expression by WB (**Figure 16A**). Full-length N6 protein of the expected size (~190 kDa in cell lysate and ~170 kDa in the

medium), as well as the excised DnaE inteins (~18 kDa), were detected only when the Npu intein set was used. Additionally, I measured the activity levels of the secreted F8 in the medium using a chromogenic assay. In the presence of full-length N6, activity levels averaged ~60 IU/dL, while single halves and the heterologous intein set exhibited little to no activity (**Figure 16B**).

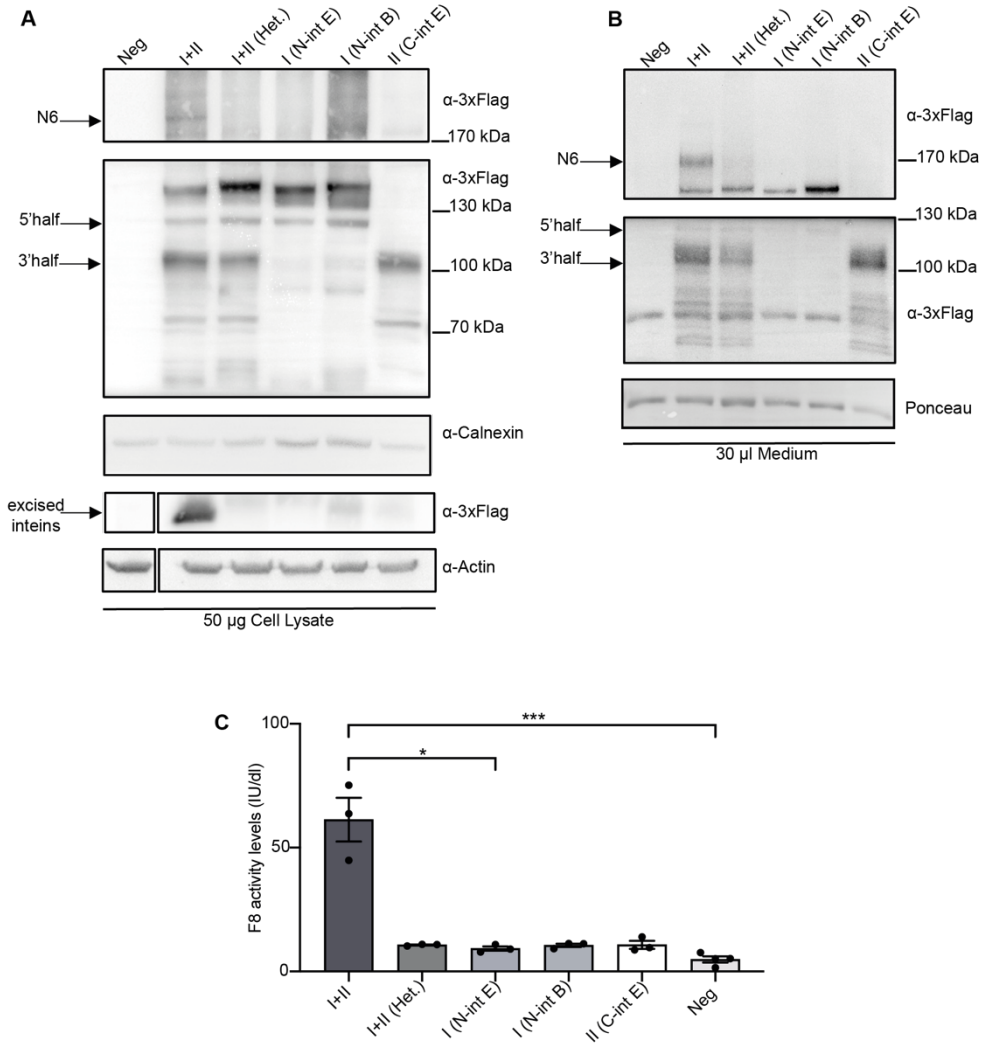
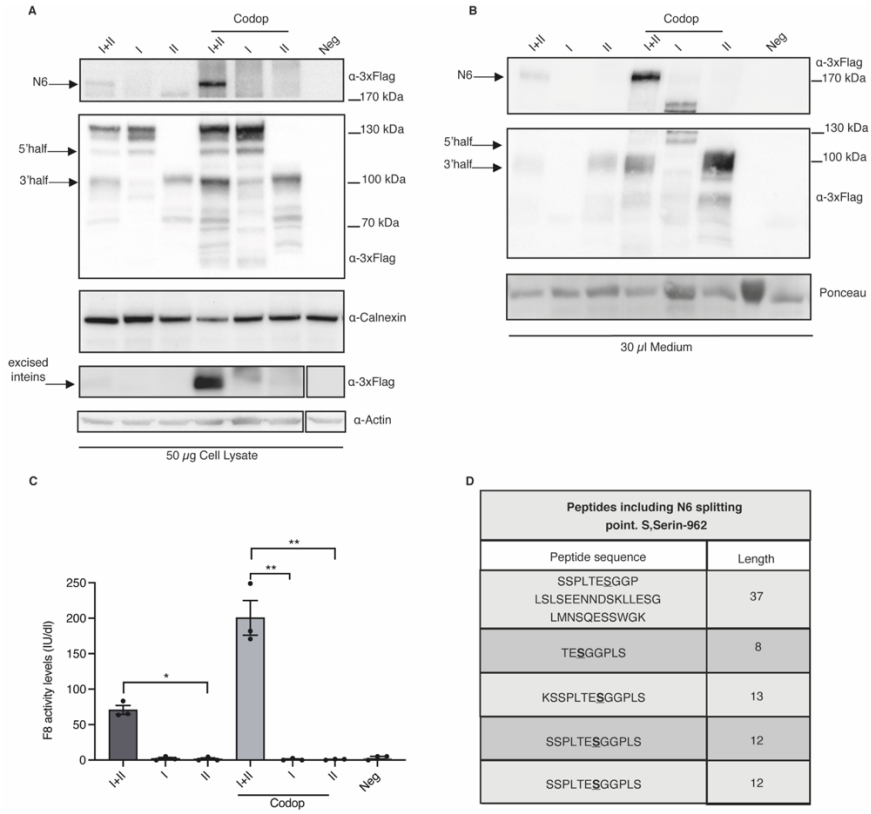


Figure 16. Evaluation of F8-N6 (N6) expression and activity with heterologous split-inteins. A) Western blot (WB) of protein lysates of HEK293 cells 72 hours post-transfection with either Npu inteins or Heterologous (N-intein DnaB + C-intein DnaE) split-inteins. I + II, N6 split-intein proteins; I+II (Het.), heterologous split-intein proteins; I, 5'N6 coding sequence (CDS)-N-DnaE protein; I (N-int B), 5'N6 CDS-N-DnaB. II, C-DnaE-3'N6 CDS protein. Excised inteins (~12kDa) are present only in the down part of the blot when I + II (Npu inteins) are provided. Arrows indicate the full-length N6 protein, single halves and excised inteins. **B)**

WB of medium of the transfected cells showing the secreted proteins. Arrows indicate the full-length N6 protein as well as single halves and excised inteins. **C)** Chromogenic assay performed on the medium of transfected cells to detect F8 activity levels reported as International Units/deciliter (IU/dl). Data are presented as mean ±SEM. Significant differences between groups were assessed using Kruskal-Wallis rank sum test Kruskal−Wallis, p value = 0.013.

N6 codon optimization enhances protein expression and activity *in vitro*

Codon optimization is an approach to increase the efficiency of protein expression. To further enhance intein efficiency, I codon-optimized the N6 intein. By transfecting the codon-optimized (CodopN6) N6 intein into HEK293 cells along with the non codon-optimized one, I observed a 4fold increase in CodopN6 intein expression and secretion by WB compared to the non-codon-optimized N6 split inteins (**Figure 17A-B**). Additionally, cells expressing CodopN6 inteins exhibited higher F8 activity levels (~200 IU/dL) than the corresponding non-codon-optimized N6 (~70 IU/dL), as assessed by chromogenic assay (**Figure 17C**). Moreover, to demonstrate that PTS results in precise CodopN6 reconstitution, I performed a liquid chromatography-mass spectrometry (LC-MS) analysis and sequenced the splitting point. The LC-MS analysis showed reconstituted CodopN6 peptides with sequences at the splitting point identical to the full-length CodopN6 encoded by a single plasmid (**Figure 17D**).



Peptides including N6 splitting point. S,Serin-962	
Peptide sequence	Length
SSPLTE<u>S</u>GGP LSLSEENNDSKLLESG LMNSQESSWGK	37
TE<u>S</u>GGPLS	8
KSSPLTE<u>S</u>GGPLS	13
SSPLTE<u>S</u>GGPLS	12
SSPLTE<u>S</u>GGPLS	12

Figure 17. Codon optimisation of the N6 split-intein improves F8 activity levels in vitro. **A**) Western blot (WB) of protein lysates of HEK293 cells 72 hpt with the AAV-N6 split- intein plasmids and with the codon-optimised set. I + II, N6 split- intein proteins; I, 5' N6 CDS-N-DnaE protein; II, C-DnaE-3'N6 CDS protein. Arrows indicate the full-length N6 protein, excised inteins and both single halves. Codop: codon-optimised. **B**) WB of medium from the transfected cells showing increased secretion of the Codop-N6 full-length protein compared to the non-codon-optimised. Arrows indicate the full-length N6 protein, excised inteins and both single halves. **C**) Chromogenic assay performed on the medium from transfected cells to measure F8 activity levels reported as International Units/deciliter (IU/dl). Data are presented as mean ±SEM. Significant differences between groups were assessed using Kruskal-Wallis test p value = 0.027.**D**) Peptides sequences obtained by LC-MS analysis which include the N6 splitting point which is correctly reconstituted; <u>S</u>: Ser 962 (n=5).

Evaluation of the structural Integrity of AAV-intein genomes

Given the positive results achieved in cells, I proceeded with the AAV vector production for: AAV-N6 intein (N6 intein), AAV-CodopN6 intein (CodopN6 intein), and the single AAV-Codon optimized F8-V3 (CodopV3) as a gold standard [59]. AAV8 vectors were generated using the small hybrid liver promoter (HLP) [120]. The genomes of the N6 intein and CodopN6 intein AAVs (3.7 kb for the 5' half and 3 kb for the 3' half) were well fitting within the AAV packaging capacity, unlike the single AAV-CodopV3 genome (5.2 kb), which exceeds the AAV cargo capacity [59]. I confirmed the packaged genomic integrity of the 3' and 5' AAV-N6 intein by alkaline Southern blot hybridization of purified vector DNA using a probe specific for the HLP promoter. The AAV-intein vector lanes showed bands of the expected molecular weight, while the AAV8-CodopV3 vector lane exhibited a heterogeneous population of truncated genomes of various sizes (**Figure 18**).

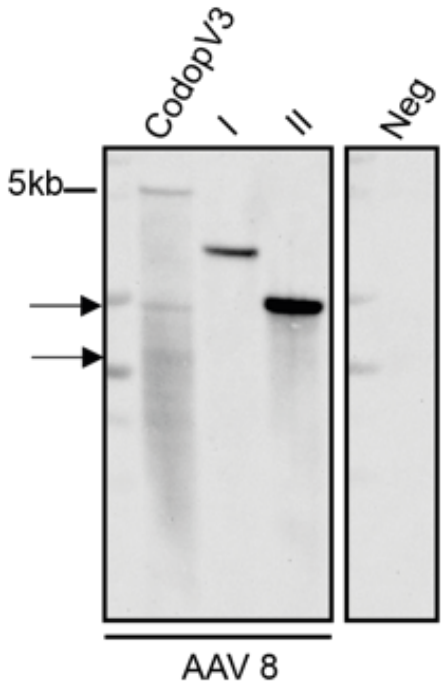


Figure 18. Southern blot analysis on AAV N6 intein vectors and on the single AAV-CodopV3. Alkaline gel Southern blot analysis of AAV DNA. AAV DNA was hybridized to a probe specific for the HLP promoter. Neg, AAV DNA treated with Dnase I; CodopV3, AAV-CodopV3; I, AAV-5'N6-N-intein; II, AAV-C-intein-3' N6.

AAV-N6 intein efficacy in HemA mice

To study the efficacy of my AAVN6 intein platform, I retro-orbitally injected adult (7-11 week-old) HemA mice (B6;129S-F8tm1Kaz; Jackson laboratory) at a dose of 5 x 10^11 genome copies (GC) per vector per animal. Specifically, HemA male mice were divided into 2 different groups of treatment: a group of mice was injected with the AAV-N6 intein while a second group was administered with the single AAV-CodopV3. Moreover, since codon optimization of N6 intein allows higher levels of F8 activity in vitro, I hypothesized that the AAV-CodopN6 intein set could achieve similar therapeutic efficacy at a lower AAV doses compared to the non-codon optimized AAV-N6 intein ones. Therefore, I administered the AAV-CodopN6 intein set at a dose of 1.5 x 10^11 genome copies per vector. I then collected blood plasma samples over 16 weeks post-vector administration at

4-week intervals and I monitored F8 activity using functional chromogenic assay. Interestingly, following AAV-N6 intein administration (whether Codop or not), plasma F8 activity reached wild-type levels (~150 IU/dL) and remained stable up to 16 weeks post-injection (wpi). These levels were comparable to those observed in animals receiving the single AAV-CodopV3 (**Figure 19A**). However, most animals (5 out of 8) treated with the single oversized AAV-CodopV3 vector showed a significant decrease or total loss of F8 activity over time. Next, to demonstrate that transduced hepatocytes expressed the full-length N6 protein, I performed WB analysis on liver lysate samples from the AAV-CodopN6 intein injected group. The analysis revealed a band of the expected molecular weight (**Figure 19B**), confirming the expression of the N6 full-length protein in the liver cells.

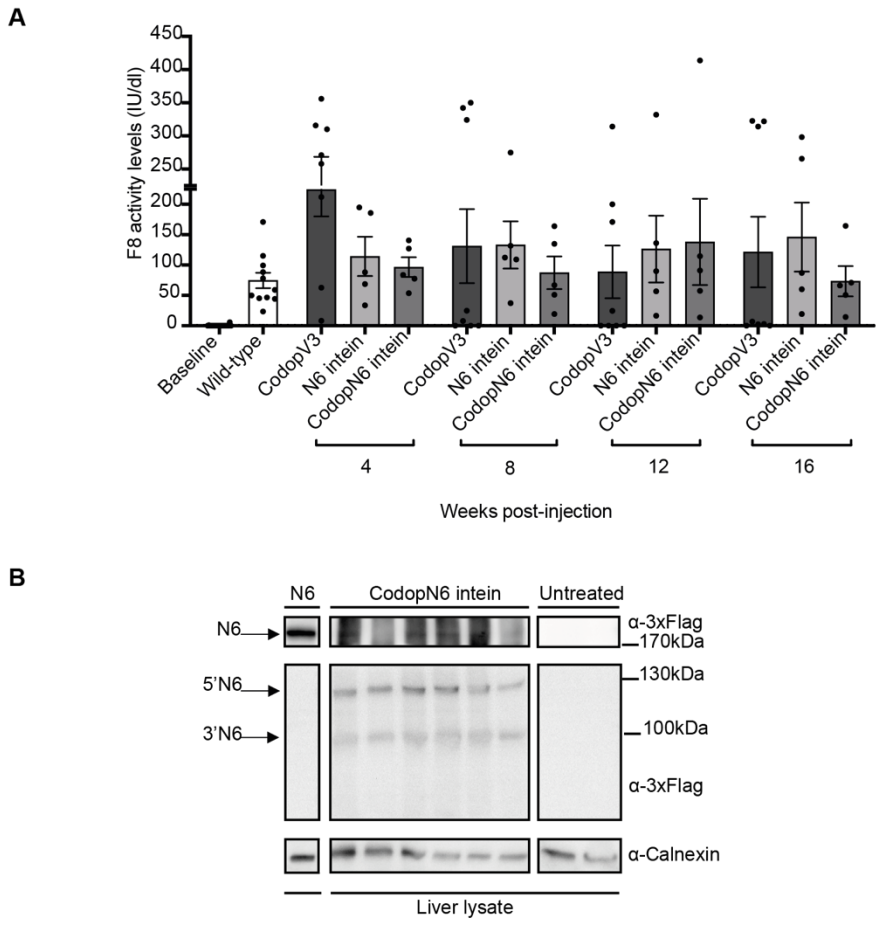


Figure 19. AAV-N6 intein result in F8 activity levels and full-length protein expression. A) Chromogenic assay performed on plasma samples to detect F8 activity in AAV-treated mice compared to controls groups. Data are presented as mean ±SEM. Statistical difference between groups as been assessed at 4weeks post-injection (wpi) with the One-way Anova p value = 0.056; at 8 wpi with the Kruskal-Wallis test p value = 0.68;at 12wpi with the Kruskal-Wallis test p value =0.45; at 16wpi with the Kruskal-Wallis test p value = 0.58. **B**) Western blot analysis of liver lysates (100 µg) from either CodopN6 intein-treated mice (n=6) or untreated haemophilic mice (n=2). A lysate from HEK293 cells transfected with the N6 full-length plasmid was used as positive control (50 µg). Arrows point at N6 full-length protein (N6), the 5' half of N6 (5'N6) and 3' half of N6 (3'N6).

AAV-N6 intein treatment improves the haemophilic phenotype in mice

To confirm that the levels of F8 activity observed in plasma samples were sufficient to ameliorate the haemophilic phenotype, I measured the activated partial thromboplastin time (aPTT) in mouse plasma 16 weeks post-injection following AAV-intein administration. aPTT was significantly decreased to normal levels in animals that received either the AAV-N6 intein or the AAV-CodopN6 intein treatment (**Figure 20A**). Furthermore, phenotype improvement in the AAV-intein treated mice was also confirmed by the tail-clip assay. However not at significant levels, all groups of AAV-treated mice showed a trend towards reduced bleeding time (**Figure 20B**) and blood loss (**Figure 20C**) compared to hemophilic untreated controls at different timepoints after AAV treatment.

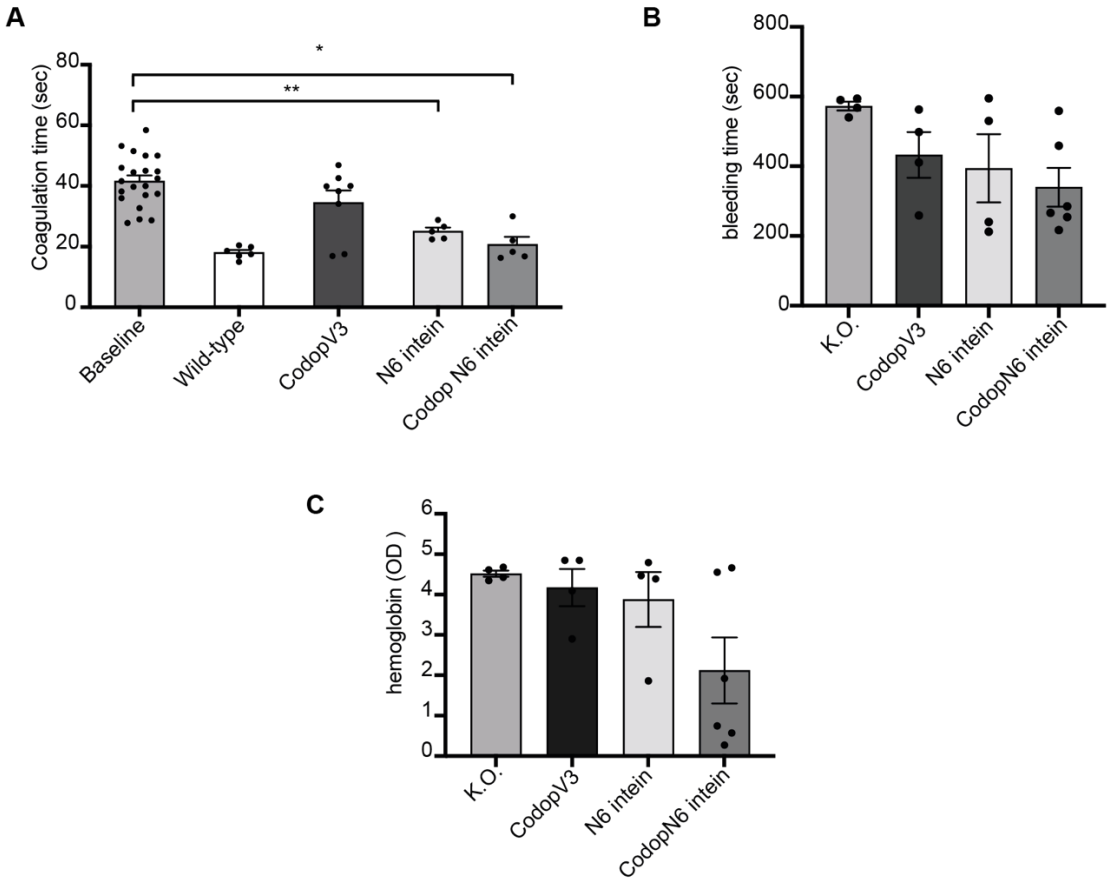


Figure 20. AAV-N6 intein improves mouse haemophilic phenotype. A) Activated partial thromboplastin time (aPTT) assay reported as Coagulation time in seconds, was performed on plasma samples both at the baseline and at the last time-point of the analysis (16 wpi). Data are presented as mean ±SEM. Significant differences between groups were assessed using the Kruskal-Wallis test. **B**) Tail-clip assay reported as bleeding time in seconds. Significant differences between groups were assessed using One-way Anova test p value is =0.11. Data are presented as mean±SEM. **C**) Haemoglobin content measured as Optical Density (OD) at 416 nm after the tail-clip assay. Significant differences between groups were assessed using Anova test.

AAV-N6 intein do not elicit anti-F8 antibodies development

One of the major drowback of treatment for HemA is the high chance of developing neutralising anti-F8 antibodies (inhibitors). To investigate this aspect, I measured anti-F8 antibodies development in AAV-treated animals: serum from animals administered with the single AAV-CodopV3 (n=8), the AAV-N6 intein (n=5) and the AAV-CodopN6 intein (n=5), were analysed by an indirect enzyme-linked immunosorbent assay (ELISA). I found that anti-F8 antibody levels were precisely inversely correlated with the F8 activity. Indeed about 60% of the mice treated with the single AAV-CodopV3 had high levels of anti-F8 antibodies and no detectable F8 activity (*P< 0.05) while the mice exhibiting high F8 activity levels (n=3) had no anti-F8 antibodies (***P<0.0008) (**Figure 21A**). On the contrary, all the mice injected with AAV-N6 intein (n=5) or with AAV-CodopN6 intein (n=5) had high F8 activity levels and no detectable anti-F8 antibodies *p< 0.05 and **p< 0.005 respectively. Since not all anti F8 antibodies are considered F8 inhibitors, I specifically investigated the possible inhibitory activity of the anti-F8 antibodies. To this aim I analysed plasma samples of mice injected with the single AAV-CodopV3 which had high levels of anti-F8 antibodies measured by indirect ELISA. By performing the Bethesda assay, I confirmed the inhibitory activity of the anti-F8 antibodies in this group of mice (**Figure 21B**). Lastly, to characterize potential differences in immune responses to F8 variants in AAV-injected mice, I analysed their serum cytokines levels both at baseline and at 4-, 8- and 16-weeks post-injection. Pro-inflammatory cytokines, including eotaxin and interferon-gamma-induced protein 10 (IP-10)[143,144] , were significantly higher in mice injected with the single AAV-CodopV3 at the different analyzed time points after AAV administration compared to baseline levels (**Figure 21C**).

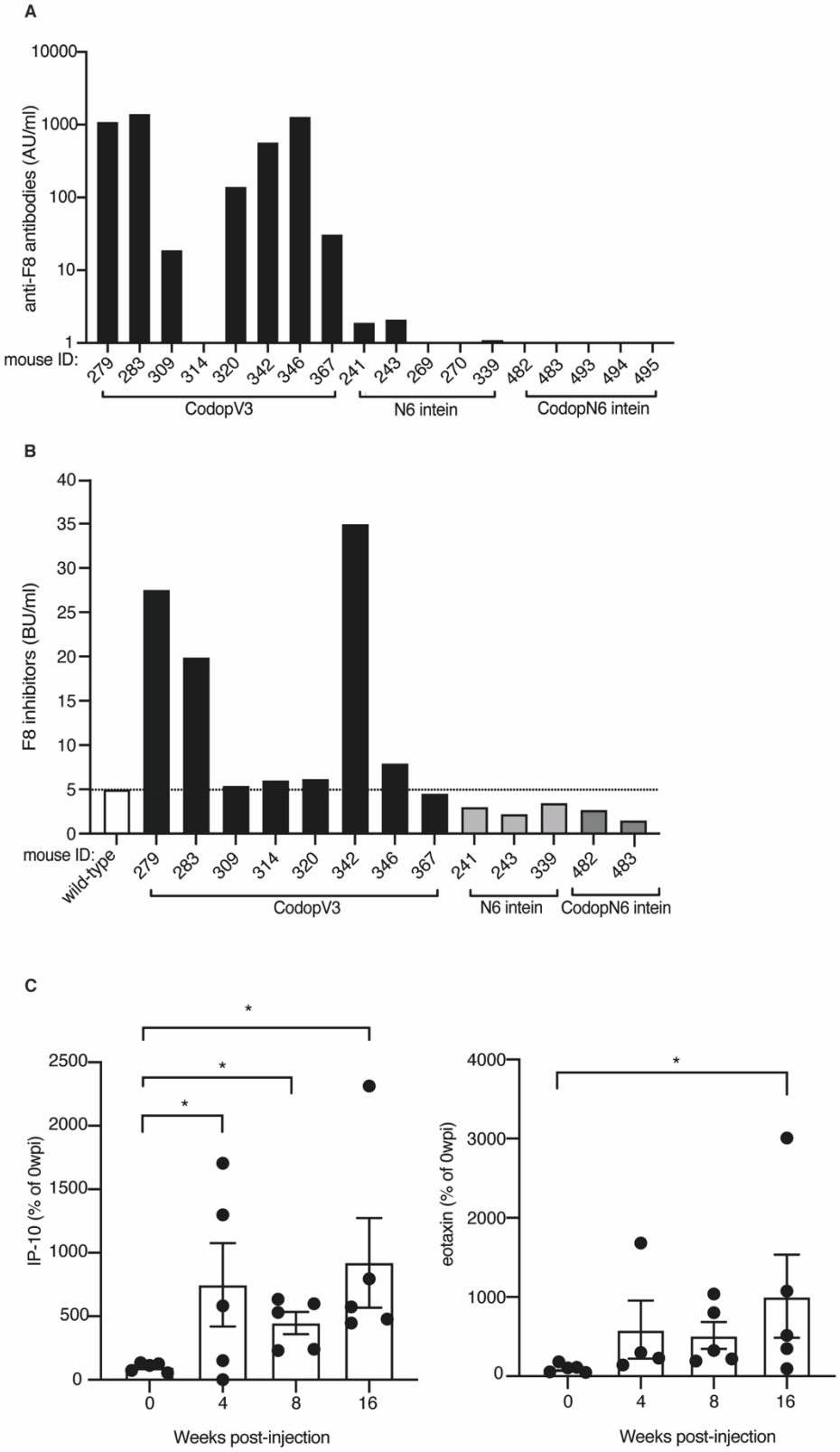


Figure 21. AAV-N6 intein administration does not elicit anti-F8 antibodies. **A**) The amount of anti-F8 antibodies analysed by indirect ELISA is reported in Arbitrary Units/milliliter (AU/ml). Each numbered bar represents a single mouse. CodopV3, AAV-CodopV3 injected group (n=8); N6 intein, AAV-N6 intein injected group (n=5); CodopN6 intein, AAV-CodopN6

intein injected group (n=5). Significant differences were assessed as follow: Paired T-test has been used for CodopV3 p value = 0.00072 (n=5) and Wilcoxon test for CodopV3 p value= 0.031 (n=3); Paired T-test has been used for N6 intein p value= 0.033 (n=5); Paired T- test has been used for CodopN6 intein p value =0.002 (n=5). **B**) Bethesda assay performed on plasma samples obtained from mice injected with AAV-CodopV3; on the plasma of three animals injected with AAV-N6 intein and the plasma of two animals injected with AAV-CodopN6 intein; plasma of a wild-type animal was used as a control. **C**) Pro-inflammatory cytokines levels were monitored over time in plasma samples obtained from the AAV-CodopV3 group of mice and reported as percentage of the baseline (time point before the injection). Significant differences were assessed using the Kruskal−Wallis test for IP-10 p value = 0,011, (n=20); for eotaxin p value = 0,036 (n=20).

Specific aim 2: The second objective of my project has focused at overcoming the loss of the therapeutic efficacy in proliferative tissues such as neonatal livers, following AAV treatment.

Liver-directed AAV-HITI platform targeting the 3' end of the mouse Albumin gene

Site-specific integration of a therapeutic donor DNA cassette into the genome ensures persistent transgene expression through cell divisions. Albumin (Alb) is recognized as a safe harbor site for gene editing applications [145]. In addition, is a highly transcriptionally active site into hepatocytes. Therefore, I investigated the feasibility of using the AAV-HITI platform to target the 3' mouse Alb (mAlb) locus. Moreover, to preserve the endogenous albumin expression after transgene integration, I evaluated the targeting of the mAlb intron 13 (**Figure 22A**). Our AAV-HITI platform relies on the co-delivery of two different AAV8 vectors. One AAV delivers the large *S. pyogenes* Cas9 (*Sp*Cas9)[148] nuclease together with a small hybrid liver-specific promoter (HLP), while the other carries the HITI donor DNA which is designed to contain a synthetic splicing acceptor signal (SAS) followed by the last m*Alb* exon (exon 14), the Thosea asigna virus 2A skipping peptide (T2A) and the promoter-less coding sequence (CDS) of a desired transgene together with a polyadenylation signal and the gRNA (or scRNA) expression cassette including the U6 promoter (**Figure 22B**). As previously described [95,146,147], the HITI donor DNA is flanked at its 5' and 3' extremities by the same gRNA sequence of the endogenous target site (m*Alb* intron 13) but in an inverted orientation. Upon CRISPR/Cas9-mediated cleavage at both the endogenous locus and at the extremities of the HITI donor DNA, the non-homologous end-joining (NHEJ) repair pathway of the cell promotes donor DNA integration at the nuclease-induced double-strand-breaks (DSBs) in the endogenous locus [95,146,147] (**Figure 22C**). Once proper integration has occurred,

the therapeutic transgene is expressed under the control of the endogenous promoter. A single fusion mRNA transcript is generated; at the protein levels, this will result in the production of both a modified albumin (Alb-2A) and a therapeutic transgene product containing a proline (P) residue at its N-terminus (**Figure 22D**).

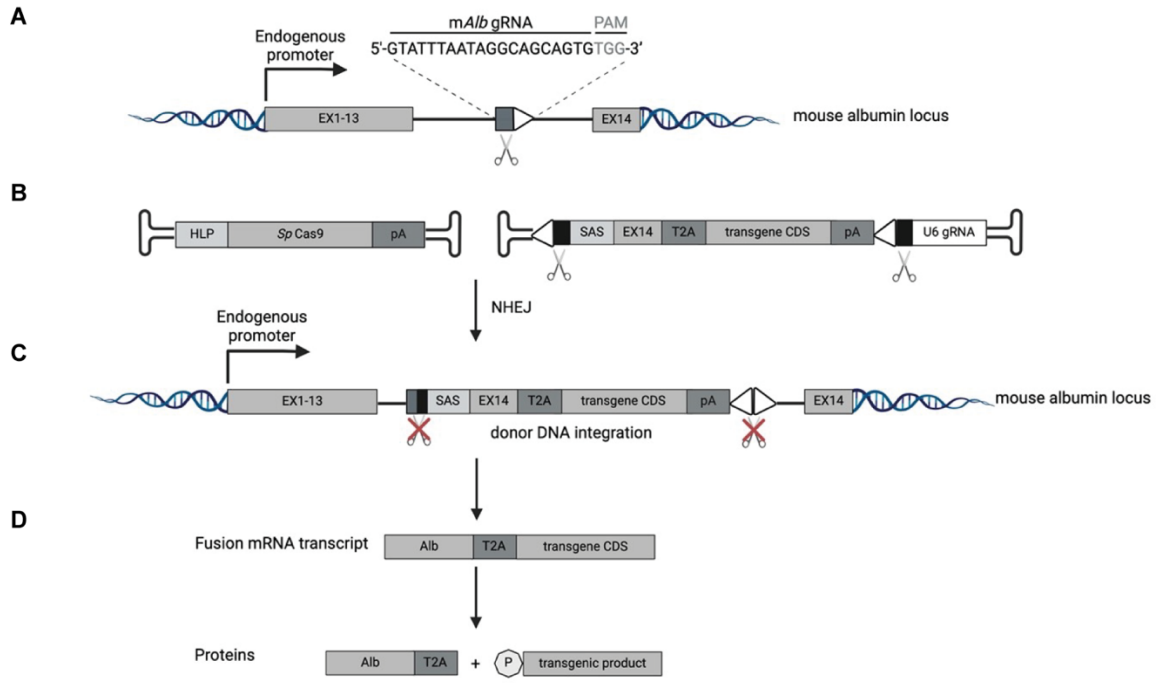


Figure 22. Schematic representation of homology-independent targeted integration (HITI) at the 3' Alb locus.A) The on-target mouse albumin locus and the gRNA sequence designed within the intron 13 are depicted. EX1-13, mouse albumin exons; EX14, last mouse albumin exon. **B)** Schematic representation of the two AAVs. One AAV carries the nuclease SpCas9 under control of the small hybrid liver-specific promoter (HLP). The other AAV contains the HITI donor DNA with the desired promoter-less transgene coding sequence (CDS). SAS, splicing acceptor signal; EX14, mouse albumin exon 14; T2A, Thosea asigna virus 2A skipping peptide; pA, polyadenylation signal (synthetic or bovine growth hormone); U6 gRNA, expression cassette for the gRNA (depicted) or the scRNA sequence; EX1-13, mouse albumin exons 1 to 13; gRNA and PAM sequences are represented with a black (inverted gRNA sequences at the extremities of the donor DNA) and grey (within the mouse albumin locus) boxes and white triangle, respectively. **C)** The non-homologous end joining (NHEJ) repair pathway of the cell leads to the integration of the cleaved donor DNA at the on-target site. **D)** Upon transgene integration a single fusion transcript is produced, and this results in the

expression of both a modified albumin (Alb2A) and a therapeutic transgene product containing a proline (P) residue at its N-terminus.

Evaluation of HITI efficiency in Hepa 1-6 cells

To set-up the platform and assess the efficiency of the gRNA, I transfected Hepa 1-6 cells with a plasmid encoding for the nuclease SpCas9 cloned in frame with the CDS of the reporter eGFP transgene under the expression of the Chicken beta-actin hybrid (CBh) promoter along with either the U6-gRNA (or -scRNA) expression cassettes. Forty-eight hours post-transfection I sorted and collected the eGFP+ve cells (transfected with the Cas9-eGFP plasmid) by performing an advanced variant of flow cytometry, fluorescence-activated cell sorting (FACS). After FACS sorting of eGFP+ve cells, I extracted the genomic DNA to evaluate the efficiency of the selected gRNA. To this aim I performed an in vitro T7 assay by PCR amplifying the region surrounding the cutting site with specific primers. TIDE analysis revealed that about 53.3% of the alleles were edited (indels) only in cells treated with the gRNA (**Figure 23**). Sanger sequencing results also showed that gRNA-treated cells were cleaved at the expected site.

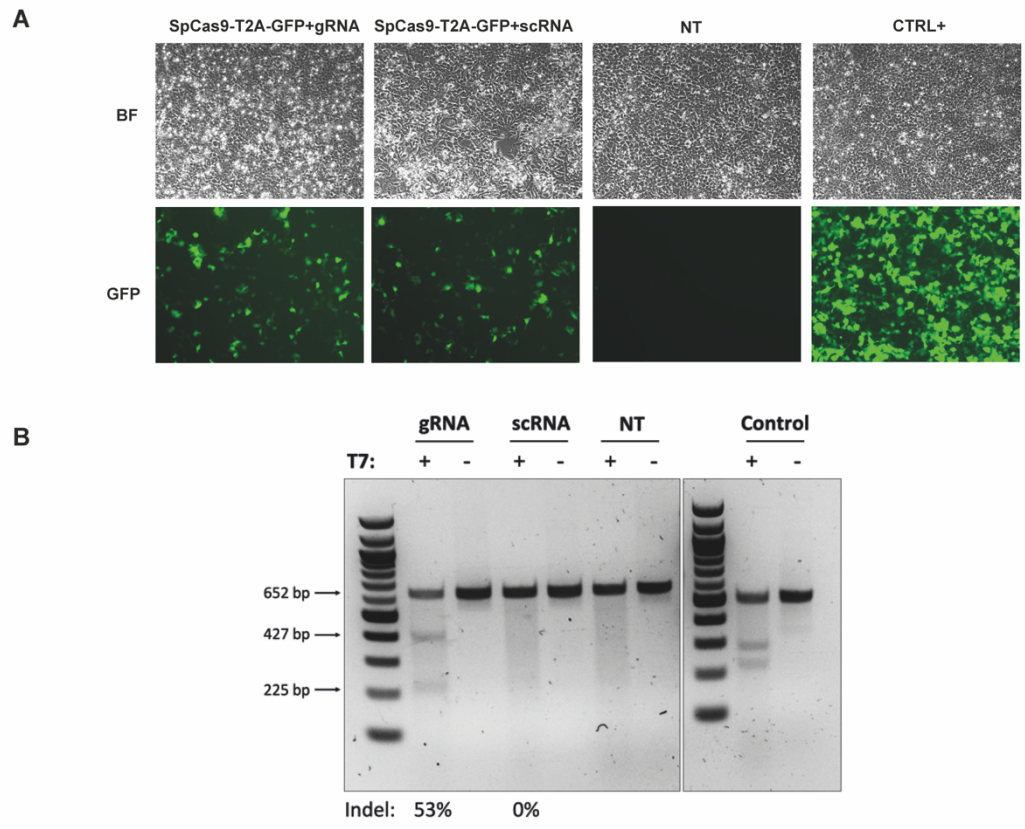


Figure 23. In vitro validation of gRNA efficiency into the 3' Alb locus. A) Fluorescence microscopy imaging of transfected Hepa 1-6 cells with a mixture of SpCas9-T2A-eGFP-gRNA (or -scRNA). NT: untreated cells; CTRL+ plasmid expressing eGFP under the CMV promoter. **B**) Analysis of indel presence by T7 endonuclease (T7) cleavage. Expected bands sizes and average indel frequency (%) are depicted (n=3). gRNA = HITI donor + SpCas9-gRNA; scRNA = HITI donor + SpCas9-scRNA; NT = non-transfected.

Given the high efficiency of the gRNA, I then designed a HITI donor DNA with the following features: i. a Synthetic Splicing Acceptor signal (SAS) to ensure the correct splicing between the endogenous locus and the integrated DNA; ii. the last mouse Alb exon (exon 14) to preserve the expression of the endogenous albumin; iii. the Thosea asigna virus 2A skipping peptide (T2A) cloned in frame with the CDS of the transgenes to get integrated. Moreover, the HITI donor DNA presents inverted gRNA sequences at both its 5' and 3' extremities. To assess HITI efficiency, I used a HITI donor DNA carrying the CDS of the reporter DsRed

transgene. For this purpose, Hepa 1-6 cells were co transfected with the HITI donor plasmid and the plasmid encoding for the nuclease linked to the eGFP fluorescent reporter (Cas9-T2A-eGFP plasmid) and analyzed by fluorescence microscopy. Seventy-two hours post- transfection, I observed an increased presence of DsRed+ve cells in the gRNA-treated group (**Figure 24A**). Quantification of the fluorescence intensity confirmed that gRNA-treated cells presented higher DsRed expression (103.2 AU) than scRNA- and then the only donor-treated cells (32.6 AU and 36.8 AU, respectively), while no significant differences in eGFP expression were observed in cells transfected with the SpCas9-T2A-eGFP plasmid with either the U6 gRNA or U6scRNA expression cassettes; **Figure 24B**).