Universitat Autònoma de Barcelona

Preclinical assessment of a gene editing approach for MNGIE

Marta Parés Casellas

Doctoral thesis

Preclinical assessment of a gene editing approach for MNGIE

Marta Parés Casellas

Thesis from Departament de Bioquímica i Biologia Molecular of Universitat Autònoma de Barcelona: PhD degree in Biochemistry, Molecular Biology and Biomedicine

This work was carried out at Gene and Cell Therapy group of Institut de Recerca Hospital Universitari Vall d'Hebron

Director

Jordi Barquinero Máñez

Tutor

Assumpció Bosch Merino

Universitat Autònoma de Barcelona

2020

The doctoral thesis titled “Preclinical assessment of a gene editing approach for MNGIE” was carried out by Marta Parés Casellas under the direction of Dr. Jordi Barquinero in the Gene and Cell Therapy group of Institut de Recerca Hospital Universitari Vall d’Hebron. The director certifies that this memory is suitable to be presented for the PhD degree in Biochemistry, Molecular Biology and Biomedicine of Universitat Autònoma de Barcelona.

This research was funded by Fondo de Investigaciones Sanitarias (FIS) of Instituto de Salud Carlos III (Ministerio de Ciencia e Innovación del Gobierno Español) through the project “Nuevas estrategias para mejorar la eficacia y reducir la inmunogenicidad en terapia génica de enfermedades hereditarias” (PI15/00172). The Fundació “la Caixa” provided the scholarship for the PhD student.

Barcelona, 16th of September of 2020.

Director: Dr. Jordi Barquinero

PhD student: Marta Parés

En primer lloc vull agrair a en Jordi, el meu director de tesi, el seu suport, orientació constant i interès al llarg de la realització de la meva tesi. Gràcies per haver confiat en mi per tal de dur a terme aquesta investigació. Els teus valuosos comentaris, revisions i suport al llarg de tot aquest temps m'han ajudat a posar en ordre les idees i estructurar el meu treball. Treballant amb tu he pogut comprovar que la teràpia gènica és un camp molt apassionant en el qual m'agradaria poder seguir vinculada. També m'agradaria agrair-te la teva proximitat, i haver estat disponible sempre que ho he necessitat.

Així mateix, vull agrair a la Sílvia la formació rebuda durant les pràctiques del màster. Has estat un dels meus referents a seguir. Això va fer que em decidís a seguir treballant en el camp d'investigació i m'animés a fer el doctorat.

Durant tots aquests anys he compartit laboratori amb molta gent diferent, alguns més temps i d'altres menys, però tots han tingut un paper rellevant en la meva vida. Molts d'aquests companys i amics no han intervingut directament en la tesi, però sí que han participat del meu dia a dia durant aquests anys. Andrea, Èlia, Paula i Bet, tot i que ja fa un temps que vau marxar, encara recordo les nostres escapades per fer un cafè i les nostres xerrades. I molta altra gent, com en Lluís, la Kelly, en Rafa, l'Eva… Tots heu contribuït a fer que sempre hi hagi hagut molt bon ambient al laboratori; ha estat un plaer compartir aquesta etapa de la meva vida amb tots vosaltres.

No em puc oblidar de mencionar a totes les estudiants de màster que he supervisat al llarg de tot aquest temps. Laura M, Laura P, Lídia, i Carla, m'ha agradat molt poder-vos ajudar durant la vostra estada. Algunes de vosaltres ja heu començat el vostre propi doctorat, i d'altres encara esteu meditant el vostre futur, però espero que totes us emporteu bons records d'aquestes pràctiques. Us desitjo molta sort en el que decidiu fer a partir d'ara.

Penélope, encara que no compartíssim laboratori sí que hem compartit tutor. Gràcies a tu el congrés de Barcelona va ser més entretingut durant els descansos. Molta sort en la teva tesi, segur que et va molt bé.

No em puc deixar de mencionar a en Jesús, l'última incorporació al laboratori. Els dinars són molt més animats des que hi ets i amb tu sempre hi ha temes de conversa. Tot i que ha sigut curt, m'ho he passat molt bé.

També vull agrair la col·laboració de la Cristina i en Salvador del IQS. M'han ajudat en l'elaboració de les nanopartícules i m'han assessorat en tot moment.

I a en Ferran, ja que sense ell una part important de la meva tesi no hauria estat possible. Et desitjo molta sort en la teva tesi també, ja que la presentarem gairebé al mateix temps. Així mateix també li vull agrair a en Ramon el seu interés i assessorament.

A més, vull agrair a tots els treballadors d'Acuitas Therapeutics® que ens han ajudat en la col·laboració científica d'elaboració de nanopartícules.

També vull agrair a en Fran i la Irene la seva col·laboració en la seqüenciació de mostres i el seu assessorament en diferents aspectes de la tesi.

A més a més, vull agrair la col·laboració de totes aquelles persones i institucions que ha fet possible la realització d'aquesta tesi. En especial, agraeixo als membres de la comissió de seguiment i a la meva tutora l'interès mostrat en l'evolució de la meva tesi. També a la Fundació “la Caixa” per haver-me atorgat la beca que m’ha permès realitzar aquest treball.

També agrair tots aquells que d'una manera o altra m'han ajudat encara que el seu nom no figuri de forma explícita en aquestes línies; sense ells tampoc hauria estat possible.

Tots vosaltres heu fet de la meva estada al VHIR una experiència única.

Per últim, vull agrair a la meva família, en especial a la meva mare, tot el suport i ànims rebuts. Tots han estat al meu costat durant la realització d'aquest treball.

ABSTRACT

MNGIE is a rare metabolic disease caused by recessive mutations in the TYMP gene, which encodes the enzyme TP. This causes a systemic accumulation of nucleosides which results in mitochondrial toxicity that is usually lethal during the first decades of life.

The goal of the present work was to achieve an efficient integration of the human *TYMP* cDNA on introns of the *Tymp* and *Alb* genes of murine hepatocytes by the coordinated action of two elements: CRISPR/Cas9 system and *TYMP* cDNA templates. These templates were designed to be integrated in a region downstream a genomic promoter, so that the *TYMP* cDNA expression will be under its control. In the case of the *Alb* locus the successful gene editing produced a hybrid Alb-hTP protein with a secretory ability. With this approach, hepatocytes would be permanently edited and the *TYMP* cDNA would be expressed long-term, thus overcoming the problem of loss of transgene expression often detected in conventional gene therapies. Two distinct strategies were tested. In the first one, CRISPR/Cas9 RNAs were delivered by nanoparticles (NPs). This allowed a high but transient expression. We tested NPs of two different sources: polymeric NPs (PNPs) of GEMAT-IQS, and lipid NPs (LNPs) of Acuitas Therapeutics®. In the second approach, CRISPR/Cas9 was delivered as DNA inside AAV vectors.

Both *in vitro* and *in vivo* results corroborated that the *TYMP* cDNA could be successfully inserted in both loci. We also assessed the presence of TYMP mRNA and functional TP protein in the edited cells. *In vivo* experimentation was performed with the mice model of MNGIE. The Thd and dUrd plasma levels of treated animals were monitored to assess the effectivity of the different genome editing approaches. The best results were obtained in animals treated with LNPs carrying the CRISPR/Cas9 RNAs and rAAV2/8 vectors carrying the DNA templates. These mice showed a consistent and stable biochemical correction that correlated with the presence of TYMP mRNA and functional TP enzyme in liver cells. Moreover, mice of the *Alb* locus gene editing group presented significant TP activity in plasma, which confirms that the hybrid enzyme can be successfully secreted and maintains its activity in plasma. However, the reduction of plasma nucleoside levels was not enough to reach the levels observed in WT mice, so there is room for optimization. Unexpectedly, in some cases genome edition was observed in the absence of CRISPR/Cas9, which suggests that the DNA templates alone could trigger the correct insertion through homologous recombination. All these results confirm that our genome editing approach is viable, although we need to increase the overall efficiency of the procedure to achieve a full biochemical correction in the murine model of MNGIE.

INDEX

1. ABBREVIATIONS

ºC	Degree Celsius
α-MEM	Alpha Minimum Essential Medium
γ-RV	Gamma-retrovirus
µg	Microgram
µl	Microliter
µM	Micromolar
µmol	Micromole
A	Adenine
Aap	Assembly-activating protein
AAT	α-1 anti-trypsin
AAV	Adeno-associated virus
ADV	Adenovirus
ADA-SCID	Adenosine deaminase SCID
Ado	Adenosine
ADP	Adenosine diphosphate
AHSCT	Allogenic hematopoietic stem cell transplantation
Alb	Albumin
ALT	Alanine transaminase
AML-12	Alpha Mouse Liver 12
AMP	Adenosine monophosphate
Amp	Ampicillin
ampR	Ampicillin resistance gene
APC	Antigen-presenting cell
ApoE	Apolipoprotein E
ASGPR	Asialoglycoprotein receptor
AST	Aspartate transaminase
ATCC	American Type Culture Collection
ATP	Adenosine triphosphate
BIR	Break-induced repair
Bp	Base pair
BSA	Bovine serum albumin
bsdR	Blasticidin resistance gene
BST	*Banc de Sang i Teixits*
C	Cytosine or carbon
C-	Negative control
C+	Positive control
CAPD	Continuous ambulatory peritoneal dialysis
CAR-T	Chimeric antigen-receptor T
Cas	CRISPR-associated protein
CBATEG	*Centre de Biotecnologia Animal i de Teràpia Gènica*
CDA	Cytidine deaminase
cDNA	Complementary DNA
CDP	Cytidine diphosphate
CFDA	China Food and Drug Administration
CjCas9	*Campylobacter jejuni* Cas9
Cm	Centimetre
CMP	Cytidine monophosphate
CMV	Cytomegalovirus
CNS	Central nervous system
CNSI	Crigler-Najjar syndrome type I
CO_2	Carbon dioxide
CRISPR	Clustered regularly interspaced short palindromic repeats
crRNA	Target-recognizing RNA
CsCl	Caesium chloride
CST	Cell Signalling Technology
Ct	Threshold cycle
Ctd	Cytidine

CTP	Cytidine triphosphate
D loop	Displacement loop
dAdo	Deoxyadenosine
dADP	Deoxyadenosine diphosphate
dAMP	Deoxyadenosine monophosphate
dATP	Deoxyadenosine triphosphate
DC	Dendritic cell
dCas9	Dead Cas9
dCDP	Deoxycytidine diphosphate
dCK	Deoxycytidine kinase
dCMP	Deoxycytidine monophosphate
dCtd	Deoxycytidine
dCTP	Deoxycytidine triphosphate
dGDP	Deoxyguanosine diphosphate
dGK	Deoxyguanosine kinase
dGMP	Deoxyguanosine monophosphate
dGTP	Deoxyguanosine triphosphate
dGuo	Deoxyguanosine
dHJ	Double Holliday junction
dKO	Double knock-out
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNDP	Deoxyribonucleoside diphosphate
dNTP	Deoxyribonucleoside triphosphate
DSB	Double strand break
dsDNA	Double-stranded DNA
DSPC	1,2-distearoyl-sn-glycero-3-phosphocholine
dsRNA	Double-stranded RNA
dTDP	Thymidine diphosphate
dTMP	Thymidine monophosphate
dTTP	Thymidine triphosphate
dUrd	Deoxyuridine
dUTP	Deoxyuridine triphosphate
E. coli	*Escherichia coli*
ECGF1	Endothelial cell growth factor 1
EETP	Erythrocyte-encapsulated TP
EFS	Elongation factor 1α short
eGFP	Enhanced green fluorescent protein
EIAV	Equine infectious anaemia virus
EMA	European Medicines Agency
ER	Endoplasmic reticulum
ERT	Enzyme replacement therapy
EtBr	Ethidium bromide
F	Forward
FAM	Fluorescein amidite
FBS	Foetal bovine serum
FDA	USA Food and Drug Administration
FIS	*Fondo de Investigaciones Sanitarias*
FIV	Feline immunodeficiency virus
FIX	Factor 9
FokI	*Flavobacterium okeanokoite* type II restriction endonuclease
G	Guanine or gram or gauge
Gc	Genome copies
GDP	Guanosine diphosphate
GEMAT	*Grup d'Enginyeria de Materials*
GMP	Guanosine monophosphate
gRNA	Guide RNA

GTP	Guanosine triphosphate
Guo	Guanosine
GVHD	Graft versus host disease
H	Histidine, hydrogen or hour
HA	Homology arm
HCC	Hepatocellular carcinoma
HDR	Homology-directed repair
HITI	Homology-independent targeted insertion
HIV	Human immunodeficiency virus
HPLC	High performance liquid chromatography
HR	Homologous recombination
HRP	Horseradish peroxidase
HSCT	Hematopoietic stem cell transplantation
HSV	Herpes simplex virus
hTP	Human TP
hTYMP	Human TYMP
IDT	Integrated DNA Technologies
IQS	*Institut Químic de Sarrià*
ITR	Inverted terminal repeat
ITS	Insulin + transferrin + sodium selenite
K	Lysine
Kan	Kanamycin
kanR	Kanamycin resistance gene
Kb	Kilobase
kDa	Kilodalton
L	Litre
L-glu	L-glutamine
LB	Lysogeny broth
LCA	Leber congenital amaurosis
LDLR	Lipoprotein receptor
LNP	Lipid NP
LPL	Lipoprotein lipase
LSEC	Liver sinusoidal endothelial cell
LTR	Long terminal repeat
LV	Lentivirus
M	Molar
MCS	Multi-cloning site
MEM	Minimum Essential Medium
MFDS	South Korea Ministry of Food and Drug Safety
Mg	Milligram
MGB	Minor groove binder
MGE	Mobile genetic element
MHC	Major histocompatibility complex
MIT	Massachusetts Institute of Technology
Ml	Millilitre
MMEJ	Microhomology-mediated end-joining
MMLV	Moloney murine leukaemia virus
MNGIE	Mitochondrial neurogastrointestinal encephalomyopathy
MOI	Multiplicity of infection
MPS	Mucopolysaccharidosis
mRNA	Messenger RNA
mtDNA	Mitochondrial DNA
mTP	Murine TP
mTymp	Murine Tymp
N	Nitrogen or sample size
NAB	Neutralizing antibody
NaCl	Sodium chloride
nCas9	Cas9 nickase
nDNA	Nuclear DNA

NDP	Ribonucleoside diphosphate
NFQ	Non-fluorescent quencher
Ng	Nanogram
NGS	Next generation sequencing
NHEJ	Non-homologous end joining
NHP	Non-human primates
NIH	National Institute of Health
NLS	Nuclear localization signal
Nm	Nanometre
NmCas9	*Neisseria meningitidis* Cas9
Nmol	Nanomole
NP	Nanoparticle
NPC	Nuclear pore complex
NUC	Nuclease lobe
O	Oxygen
OLT	Orthotopic liver transplant
OM-pBAE	Oligopeptide end-modified poly-(β-amino ester)
Ori	Origin of replication
P	Phosphate
P/S	Penicillin and streptomycin
pA	Polyadenylation signal
PAM	Protospacer adjacent motif
PAMP	Pathogen-associated molecular pattern
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PD-ECGF	Platelet-derived endothelial cell growth factor
pegRNA	Prime editing gRNA
PEG	Poly(ethylene glycol)
Pi	Inorganic phosphate
PLA2	Phospholipase A2
PMSF	Phenylmethylsulfonyl fluoride
PNP	Polymeric NP
Poly-A	Poly-adenosines
Pre-mRNA	Precursor mRNA
PRR	Pattern recognition receptors
R	Reverse
R&D	Research and Diagnostic Systems
rAAV	Recombinant AAV
REC	α-helical recognition lobe
RNA	Ribonucleic acid
RNase	Ribonuclease
RNP	Ribonucleoprotein
RNR	Ribonucleotide reductase
RP	Retinitis pigmentosa
Rpm	Revolutions per minute
RT	Room temperature
RV	Retrovirus
S	Sulphur or second
S. pyogenes	*Streptococcus pyogenes*
SA	Splice acceptor
SaCas9	*Staphylococcus aureus* Cas9
SAP	Shrimp alkaline phosphatase
scAAV	Self-complementary AAV
SCD	Sickle cell disease
SCID	Severe combined immunodeficiency
SDS	Sodium dodecyl sulfate
SDSA	Synthesis-dependent strand annealing
SEM	Standard error of the mean
SIN	Self-inactivating
SpCas9	*Streptococcus pyogenes* Cas9

SSA	Single-strand annealing
SSB	Single-strand break
ssDNA	Single-stranded DNA
ssRNA	Single-stranded RNA
SV40	Simian virus 40
T	Thymine
TALE	Transcription activator-like effector
TALEN	TALE nuclease
TBG	Thyroxine-binding globulin
Thd	Thymidine
Thy	Thymine
TK1	Thymidine kinase 1
TK2	Thymidine kinase 2
TLR	Toll-like receptor
TP	Thymidine phosphorylase protein
tracrRNA	Trans-activating RNA
TS	Thymidylate synthase
U	Uracil or enzymatic activity units
UAB	*Universitat Autònoma de Barcelona*
UAT	*Unitat d'Alta Tecnologia*
Ub	Ubiquitin
UDP	Uridine diphosphate
UMP	Uridine monophosphate
UPP1	Uridine phosphorylase 1
UPP2	Uridine phosphorylase 2
UPV	Viral Vector Production Unit
URL	*Universitat Ramon Llull*
USA	United States of America
UTP	Uridine triphosphate
UTR	Untranslated region
V	Volt
VHIR	*Institut de Recerca de Vall d'Hebron*
WES	Whole exome sequencing
WGS	Whole genome sequencing
WT	Wild-type
x **g**	Times gravity
X-SCID	X-linked SCID
ZFN	Zinc finger nuclease

2. INTRODUCTION

2.1. Mitochondrial neurogastrointestinal encephalomyopathy (MNGIE)

MNGIE is rare progressive and degenerative disease with high morbidity and poor prognosis. It has an estimated prevalence of 1-9 cases per million population world-wide [1]. But this disease is heavily underdiagnosed, so it is not possible to establish a reliable prevalence [2]. The reported mean age at onset is 20 years [3,4], although this may not be accurate due to the diagnosis delay. In about 60% of the cases the symptoms begin before age of 20 years. The earliest reported age of onset is 5 months. The mean age of death is around 35 years [4]. MNGIE patients have diverse ethnic origins, and both genders are equally affected [5].

This disease was first described in 1976. A study reported a man with cachexia, ptosis, dysphagia, myopathy and ophthalmoplegia. Histology results revealed mitochondrial abnormalities in skeletal muscle and liver cells. The term "Congenital Oculoskeletal Myopathy" was proposed to describe the disorder [6]. From then on, similar cases were reported in literature. Particularly, a study reported leukoencephalopathy in a patient with a history of gastrointestinal symptoms, and extraocular and skeletal myopathy [7]. In 1994 a scientific group systematically reviewed all reported cases and established the current nomenclature [3]. Finally, in 1999 it was determined that the deficiency of thymidine phosphorylase (TP) was the molecular cause of the disease [8].

2.1.1. *TYMP* gene and mRNA

MNGIE is an autosomal recessive disease caused by pathogenic mutations in the *TYMP* gene, which encodes the cytoplasmic protein TP. The gene is located in the chromosomal locus 22q13.32-qter [8,9] , has a length of 4,3 kilobases (kb) and comprises 10 exons. There are 5 known TYMP mRNA variants. Four of them present alternative splicing at 5' untranslated region (UTR), but they encode for the same protein isoform. However, the fifth variant presents an additional alternative splicing at the 3' coding region, so it encodes a different protein isoform that has an additional segment in the C-terminal region.

Patients are either homozygous or compound heterozygous for *TYMP* mutations [10]. Various pathogenic mutations have been reported to date, including insertions [11], deletions [12], single nucleotide insertions/deletions [8,12,13] point mutations [8,13,14,15], splice-site alterations [8,12,16,17] and frameshift mutations [18]. Additionally, some non-pathogenic polymorphisms have been described [8,19].

2.1.2. TP enzyme

The protein is composed of two subunit homodimers, each with a molecular weight of 47 kilodaltons (kDa). It is initially synthesized as a polypeptide precursor (482 amino acids), which undergoes a proteolytic processing in the N-terminus to become functional [20]. This

enzyme was first known as "platelet-derived endothelial cell growth factor" (PD-ECGF or ECGF1) because it was mistakenly identified as a growth factor abundant in platelets [21,22].

TP catalyses the phosphorolysis of the pyrimidine nucleosides: thymidine (Thd) and deoxyuridine (dUrd). These molecules are converted to their corresponding nitrogenous bases, thymine (Thy) and uracil (U), and deoxyribose-1-phosphate is generated as a subproduct [8,23,24,25] (Fig. 1). This reaction is thermodynamically reversible, but the forward reaction is predominant in physiological conditions [26]. In addition, this protein is involved in many other cellular processes. It has been reported that TP also functions as: inhibitor of apoptosis [27], potent pro-thrombotic activator in platelets [22], inhibitor of glial cells and neurotrophic effector on cortical cells [28].

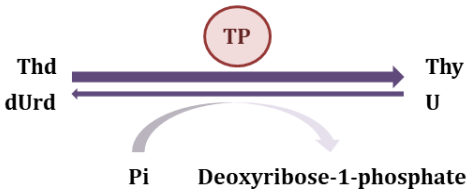

Figure 1. **TP enzymatic reaction.** Pi: inorganic phosphate.

In humans, TP is expressed in the digestive tract, liver, lungs, spleen, bladder, blood cells (platelets, stromal cells, macrophages, reticulocytes, and peripheral lymphocytes), brain, peripheral nerves and autonomic nerves. In contrast, TP is expressed at low levels in skeletal muscle, and is lacking in kidneys, aorta, gallbladder and adipose tissue [25]. The nucleosides can diffuse through cell membranes, so blood cells and tissues rich in TP regulate the extracellular and intracellular levels of Thd and dUrd of all tissues, even the ones lacking endogenous TP expression [8,10,29,30,31,32].

2.1.3. Nucleoside and nucleotide metabolism

A nucleoside is a nitrogenous base covalently attached to a sugar (pentose). There are two types of nitrogenous bases: purines (adenine and guanine) and pyrimidines (thymine, cytosine and uracil). When the pentose is a ribose, they are called ribonucleosides; and when it is a deoxyribose, they are called deoxyribonucleosides. Nucleic acids are linear sequences of nucleosides linked together by phosphodiester bonds. Each unit in these chains (nucleoside + phosphate group) is called nucleotide. The ribonucleotides are the monomeric units that compose RNA (ribonucleic acid), while the deoxyribonucleotides compose DNA (deoxyribonucleic acid) (Fig. 2, Table 1).

Nucleosides that are not part of DNA/RNA chains move freely inside cells and can be phosphorylated by different kinases. Up until three phosphate groups can be added, resulting in nucleosides mono-, di- or triphosphate [33] (Fig. 2, Table 1). The terms nucleotide and nucleoside monophosphate can be confused, since they represent the same chemical structure. When this compound is part of a DNA/RNA chain we will refer it as nucleotide, and when it is a single free molecule we will refer it as nucleoside monophosphate.

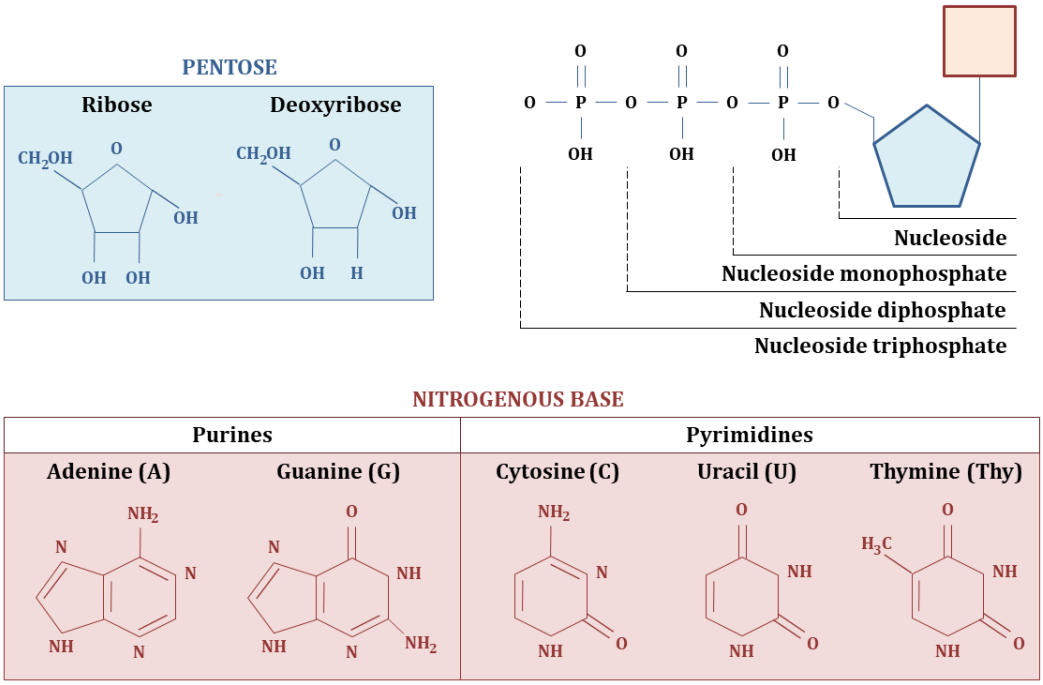

Figure 2. **Biochemical structure of nucleosides.** N: nitrogen; H: hydrogen; O: oxygen; C: carbon; P: phosphate.

DNA polymerases are the enzymes that perform the phosphodiester bond between deoxyribonucleotides during DNA replication. They carry out the hydrolytic cleavage of deoxyribonucleoside triphosphates (dNTPs) and link them to the chain. Diphosphates are the subproduct of this reaction. The deoxyribonucleotides needed for DNA replication can be synthetized by two different pathways:

- ***De novo* synthesis**: The deoxyribonucleotides are synthesized from metabolic precursors, which are different depending on the type of nitrogenous base. After some catalytic reactions, ribonucleoside diphosphates (NDPs) are obtained. Then they are catalysed by a ribonucleotide reductase (RNR) and become deoxyribonucleoside diphosphates (dNDPs). Finally, they are phosphorylated to obtain dNTPs, which can be incorporated in DNA chains as deoxyribonucleotides by DNA polymerases [33]. This pathway is regulated by the cell cycle, so it is mainly active in dividing cells [33,34,35].

- **Salvage pathway**: The deoxyribonucleotides are obtained from recycling deoxyribonucleosides that come from diet or intracellular DNA/RNA degradation. Various kinases are responsible of the phosphorylation of deoxyribonucleosides to generate dNTPs [36,37] (Table 2). Finally, DNA polymerases convert dNTPs to deoxyribonucleotides [33].

Nucleic acid	Type of nitrogenous base	Nitrogenous base	Nucleoside	Phosphorylated forms
RNA	Purines	Adenine (A)	Adenosine (Ado)	Adenosine Monophosphate (AMP) Adenosine Diphosphate (ADP) Adenosine Triphosphate (ATP)
		Guanine (G)	Guanosine (Guo)	Guanosine Monophosphate (GMP) Guanosine Diphosphate (GDP) Guanosine Triphosphate (GTP)
	Pyrimidines	Cytosine (C)	Cytidine (Ctd)	Cytidine Monophosphate (CMP) Cytidine Diphosphate (CDP) Cytidine Triphosphate (CTP)
		Uracil (U)	Uridine (Urd)	Uridine Monophosphate (UMP) Uridine Diphosphate (UDP) Uridine Triphosphate (UTP)
DNA	Purines	Adenine (A)	Deoxyadenosine (dAdo)	Deoxyadenosine Monophosphate (dAMP) Deoxyadenosine Diphosphate (dADP) Deoxyadenosine Triphosphate (dATP)
		Guanine (G)	Deoxyguanosine (dGuo)	Deoxyguanosine Monophosphate (dGMP) Deoxyguanosine Diphosphate (dGDP) Deoxyguanosine Triphosphate (dGTP)
	Pyrimidines	Cytosine (C)	Deoxycytidine (dCtd)	Deoxycytidine Monophosphate (dCMP) Deoxycytidine Diphosphate (dCDP) Deoxycytidine Triphosphate (dCTP)
		Thymine (T or Thy)	Thymidine or Deoxythymidine (Thd or dThd)	Thymidine Monophosphate (dTMP) Thymidine Diphosphate (dTDP) Thymidine Triphosphate (dTTP)

Table 1. Classification of nucleosides

The levels of the four dNTPs (dATP, dCTP, dGTP, dTTP) are strictly regulated. They have to be present at certain levels during DNA replication to avoid mutations. This regulation is important in the nucleus, where the nuclear DNA (nDNA) resides; and in the mitochondria, where the mitochondrial DNA (mtDNA) resides [33].

Mitochondrial dNTPs are either imported from the cytosol through specific transporters, or synthesized *in situ* through the salvage pathway [33,37]. While in proliferative cells the two

pathways are active, quiescent cells (muscle and neurons) rely basically on the salvage pathway. Since these cells do not divide, the nDNA does not need incorporation of new deoxyribonucleotides and the cytosolic *de novo* synthesis pathway is downregulated [33,34]. But mitochondria are continuously replicating, even in non-proliferative cells, so they need a constant supply of deoxyribonucleotides to maintain the mtDNA.

Enzyme	Target deoxyribonucleoside	Location
Thymidine kinase 1 (TK1)	Thd, dUrd	Cytosol
Thymidine kinase 2 (TK2)	Thd, dUrd, dCtd	Mitochondria
Deoxycytidine kinase (dCK)	dCtd, dGuo, dAdo	Cytosol
Deoxyguanosine kinase (dGK)	dGuo, dAdo	Mitochondria

Table 2. Enzymes responsible of the first step of the salvage pathway.

Pyrimidine metabolism

The enzyme cytidine deaminase (CDA) can convert dCtd to dUrd. Either way, all the pyrimidine deoxyribonucleosides (Thd, dUrd, dCtd) can be phosphorylated by TK1, TK2 or dCK and converted to dTMP, dUMP and dCMP, respectively. At this point, dUMP can either be further phosphorylated by kinases to obtain deoxyuridine triphosphate (dUTP), or it can be converted to dTMP by the enzyme thymidylate synthase (TS). dTMP and dCMP are chain phosphorylated by kinases to obtain dTTP and dCTP, respectively. Finally, DNA polymerases can use dUTP, dTTP and dCTP to incorporate them as nucleotides [26,33] .

TK1 is located in the cytosol and is mainly active in the phase S of the cell cycle, so it is only relevant in proliferative tissues [34,37]. dCK is also located in the cytosol, but its expression is not dependant of the cell cycle. Meanwhile, TK2 resides in mitochondria and is constitutively expressed [37] (Table 2). TS is only expressed in the cytosol [26].

dUrd is a deoxyribonucleoside, but in physiological conditions it is not present in DNA. However, DNA polymerases cannot distinguish between dTTP and dUTP because both have affinity for dAdo, so they can incorporate dUrd by mistake. Cells avoid this by maintaining a high ratio dTTP/dUTP, so there is little dUTP available to compete with dTTP. TS role is to maintain this ratio by converting dUMP to dTMP [26]. And even if some dUrd is incorporated, there are specific enzymes that correct these errors [26,30,33,38].

There is also a relationship between dTTP and dCTP pools. Both Thd and dCtd are phosphorylated by TK2 to obtain dTMP and dCMP, so both substrates compete with each other. But dCtd is a relatively inefficient competitor, so in presence of both, TK2 has propensity to catalyse Thd over dCtd. Moreover, their triphosphate forms also act as competitive inhibitors, meaning that dTTP inhibits the phosphorylation of dCtd, while dCTP inhibits the phosphorylation of Thd [37].

All these enzymatic pathways maintain the optimal pyrimidine deoxyribonucleoside ratios for the correct DNA replication in physiological conditions (Fig. 3).

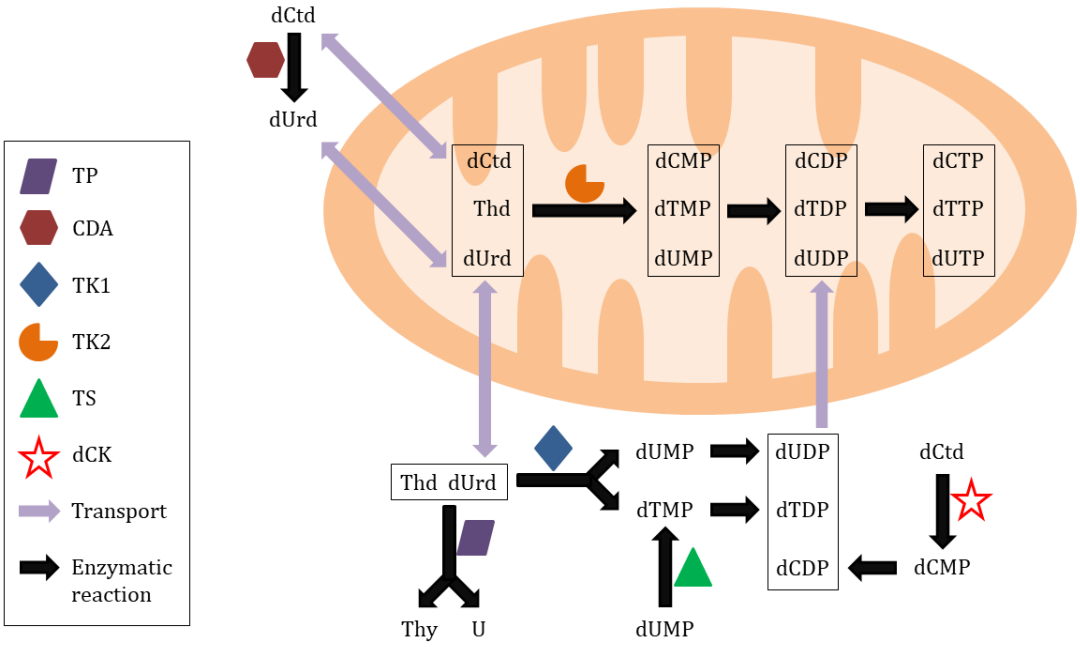

Figure 3. Schematic representation of the pyrimidine deoxyribonucleoside metabolism in mitochondria.

2.1.4. Molecular genetic pathogenesis

Thd and dUrd can either be catabolized by TP or phosphorylated by TK1 and TK2. When Thd and dUrd levels rise due to TP deficiency, the phosphorylation pathway is overly activated. Consequently, there is an excess of dTTP and dUTP, and an indirect reduction of dCTP levels, so there is an imbalance in mitochondrial dNTP pools [39,40]. So, although TP is not directly involved in the deoxyribonucleotide synthesis, its deficiency has an impact in dNTP pools.

In proliferative cells, the mitochondrial dNTPs imbalance is less pronounced because it can be compensated by the *de novo* synthesis of the other dNTPs. But in quiescent cells only the salvage pathway controls the mitochondrial dNTP pools, so the imbalance is more pronounced. That explains why the tissues most affected in MNGIE are predominantly post-mitotic [8].

This dNTPs imbalance has a mutagenic effect in mtDNA in various tissues over time. The overrepresentation of dTTP and dUTP during mtDNA replication causes an accumulation of defects (deletions and duplications), while the lack of dCTP causes the depletion of mtDNA [3,13,29,39,40,41,42,43,44]. There appear to be hotspots for mutations in mtDNA, such as the *NDS* gene, which is prone to multiple deletions [45]. Mitochondrial dysfunction causes oxidative phosphorylation impairment, which is deleterious for cells. Throughout an

individual's life the mtDNA continues to replicate, accumulating mutations over time. That explains the progressive nature of this disease [3,13].

TS is not present in mitochondria, so dUMP molecules can only be further phosphorylated to dUTP. This affects the dTTP/dUTP ratio and favours the incorporation of dUrd into mtDNA [26,39]. Cells try to repair the errors, but the cumulative effects of reparation compromise mtDNA stability [46].

Excess Thd monopolizes the enzyme TK2, so the generation of dCMP is heavily decreased. Moreover, the consequent high levels of dTTP further inhibit the phosphorylation of dCtd [37]. Additionally, it is known that the binding of dTTP to the enzyme RNR causes the specific inhibition of the conversion of CDP to dCDP, which also contributes to the dCTP deficiency [47]. All these evidences conclude that high levels of dTTP cause a depletion of dCTP. Since the mtDNA replication is limited by the availability of the dNTP present at the lowest concentration, dCTP depletion causes the reduction of mtDNA replication rate. In conclusion, mtDNA depletion is caused by dCTP deficiency rather than dTTP excess [43,47].

The fact that this imbalance damages the mtDNA but not the nuclear DNA is due to several factors. First, the mtDNA has a limited capacity to repair damage and lacks histone protection, so it is more vulnerable [48]. Second, the mitochondrial dNTP pool is sequestered within the mitochondria and regulated independently, so the imbalance is more pronounced [48,49,50]. Lastly, the mtDNA is more dependent on the pyrimidine salvage pathway than nuclear DNA, so mtDNA is more vulnerable to the toxic effects of excess Thd and dUrd [48,50].

In MNGIE patients, symptoms often appear when a threshold level of mutant mtDNA is reached. That usually happens when 80-90% of the total mitochondria are affected [13]. The mitochondria have a heteroplasmic nature, meaning that there are two or more mtDNA genotypes within the same cell. This phenomenon, coupled with the threshold effect, could very likely contribute to the heterogeneity of phenotype and age of clinical onset of MNGIE [2]. This is a problematic aspect because specific *TYMP* mutations do not necessarily correlate with a distinct phenotype, meaning that in most cases it is not possible to anticipate the disease severity [10,12]. It is also possible that environmental factors contribute to the severity of symptoms [4,19]. Although there is no clear correlation between genotype and phenotype, there is correlation between biochemical and clinical phenotype. It has been reported that certain mutations are associated with TP forms that retain residual activity, and that correlates with milder symptoms and late onset [51].

2.1.5. Clinical characteristics

MNGIE is a systemic disease with multiple phenotypic manifestations (Table 3), but the gastrointestinal, nervous and muscular systems are the most affected. The order in which manifestations appear is unpredictable. Before the overt clinical onset, the patients are usually healthy but with a long history of low weight, mild gastrointestinal symptoms or subtle fatigability [4]. Common causes of death include: aspiration pneumonia, intestinal

perforation, peritonitis, malnutrition, metabolic acidosis, and complications related to intestinal bacteria overgrowth [4,12].

Features	Frequency	Symptoms and signs
Gastrointestinal	+++	Nausea, emesis, dysphagia, gastroparesis, cachexia, weight loss, malnutrition
	++	Intestinal pseudo-obstruction, constipation, abdominal pain, borborygmi, diarrhoea, oesophageal varices, diverticulosis, intestinal perforation, peritonitis, hepatic steatosis
	+	Megacolon, hepatomegaly, cirrhosis
Neurological	+++	Peripheral neuropathy, leukoencephalopathy
	++	Hearing loss, paraesthesia, lower limb weakness
	+	Seizures, migraine, anxiety, depression, cognitive dysfunction, dementia, mental retardation, memory loss, ataxia, trigeminal neuralgia
Neuro-ophthalmic	+++	Ophthalmoplegia, ophthalmoparesis, ptosis
	+	Glaucoma, pigmentary retinopathy
Endocrine and metabolic	++	Diabetes, hyperlipidemia, hyperglyceridemia
	+	Hypergonadotropic hypogonadism, pancreatic insufficiency
Muscular	++	Myopathy, red ragged fibres, reduced muscular mass
Developmental	++	Short stature, slender and frail physique
Cardiac	+	Supraventricular tachycardia, ventricular hypertrophy, mitral valve prolapse, long QT
Reproductive	+	Ovarian failure, erectile dysfunction, amenorrhea
Dermatological	+	Psoriasis
Haematological	+	Anaemia

Table 3. Clinical features reported in MNGIE patients. The symptoms/signs are classified as major diagnostic features (+++), common clinical presentations (++) or sporadic features (+) [2,4,3,8,10,12,52].

2.1.6. Biochemical characteristics

A deficiency of TP enzymatic activity causes the accumulation of its substrates. Since Thd and dUrd can cross cell membranes via specific transporters, the nucleoside imbalance is ubiquitous. There is an accumulation of Thd and dUrd in plasma, urine and all tissues, even the ones clinically unaffected [31].

Plasma levels of deoxyribonucleosides in healthy individuals are undetectable, but in MNGIE patients the levels can reach 3,9-17,7 μmol/l for Thd, and 5,5-24,4 μmol/l for dUrd. That is an increase of 70-fold for Thd and more than 100-fold for dUrd [26]. In tissues like the small intestine, liver and kidneys, levels of 38-1532 nmol/g protein for Thd, and 32-728 nmol/g protein for dUrd have been reported [31].

The presence of these nucleotides in urine suggests that they are excreted, but only a 20% of clearance is observed. A possible explanation is that nucleosides are reabsorbed by renal tubules [12,29,30].

The quantification of TP enzymatic activity is another biochemical indicator. An activity of <10% results in severe MNGIE, while a residual activity of 10-20% produces a less severe and late onset of MNGIE [5,53]. The parents of an affected individual are obligate carriers of one *TYMP* pathogenic variant (heterozygotes), but they are not at risk of developing the disorder. Carriers have 26-35% of residual TP activity and undetectable levels of plasma Thd and dUrd [8,19,29,51]. This activity value approximates the expected 25% activity of homodimeric enzymes such as TP, because the combination of a mutated subunit and a normal subunit would produce a protein without activity. The activity value of carriers has therapeutic implications because it indicates that >26% of residual TP activity is sufficient to eliminate circulating Thd and dUrd [5].

2.1.7. Diagnosis

It is important to establish an early diagnosis, before the development of cachexia and irreversible tissue damage. In a study of the natural history of French MNGIE patients, it was determined that the initial symptoms are frequently gastrointestinal. This study reported that the clinical signs were not clear, so differential diagnoses were abundant and MNGIE was rarely diagnosed (24%). When the neurological symptoms were present the diagnosis was more accurate (56%), but at that point the cachexia was often severe [54]. Clinicians are rarely familiar with this condition, so patients typically undergo referral to different specialists before a diagnosis is achieved, and they are frequently misdiagnosed. All these factors lead to diagnostic delays of 5-10 years, and by that time the symptoms are quite severe and patients have a worse prognosis [2,4,19,32,53,52,55,56,57].

Diffuse leukoencephalopathy, a major MNGIE feature, can be observed by brain magnetic resonance imaging. The presence of this feature helps discriminate between MNGIE and other diseases with similar symptoms. Additionally, neurogenic and myogenic abnormalities are commonly detected by electromyography [12,55].

When there is a suspicion of MNGIE based on clinical examination, the molecular diagnosis can be established by one of the following tests:

- **Nucleoside quantification**: Detection of elevated plasma and urine concentrations of Thd and dUrd [19,29]. In plasma, an increase of >3 μmol/l of Thd, or >5 μmol/l of dUrd is sufficient to make the diagnosis. These nucleosides are undetectable in healthy individuals [12,19].

- **Enzymatic activity testing**: Detection of severely reduced levels of TP enzyme activity in buffy coat leukocytes. When patients show little (<10%) or no activity, it is safe to diagnose them with MNGIE [8,19,29].
- **Molecular genetic testing**: Detection of biallelic pathogenic variants in the *TYMP* gene. The first choice to detect mutations is Sanger sequencing, but sometimes it is necessary the use of other techniques such as next generation sequencing (NGS), quantitative PCR, Southern blot, gene panels, whole exome sequencing (WES), whole genome sequencing (WGS) or mtDNA studies [2,3,10,12]. At minimum, all exons with flanking intronic sequences must be sequenced. But given the diversity of *TYMP* mutations, it is possible to miss alterations in the promoter and non-flanking intronic regions [8]. Moreover, additional tests are needed to confirm the pathogenicity of a newfound polymorphism. For all these reasons, the molecular genetic testing is better for diagnosis confirmation, when there are already evidences of TP enzymatic activity alterations [19].

Since TP activity and nucleoside levels can be routinely measured in blood samples, the inclusion of MNGIE in newborn screening programs should be considered [58].

2.1.8. Treatment

Current management

MNGIE is one of the many diseases that still have no cure, so the current management is primarily supportive and requires the coordinated effort of different clinical specialities. The goal is to improve quality life and avoid complications that cause morbidity and mortality [2,59].

There are some symptoms that can be stopped with pharmacological treatment: drugs to stop nausea and emesis; analgesics; bowel motility stimulant drugs; antibiotics for intestinal bacterial overgrowth; and drugs for neuropathic symptoms [55,60].

Other ways to support the patients involve: doing physical and occupational therapy to help preserve mobility; attending to swallowing difficulties and airway protection; or using nutritional support (bolus feeding, gastrostomy tube placement, parenteral nutrition) [55,61]. Some cases require a specialized schooling arrangement.

Investigational therapies

There are many therapeutic approaches currently being investigated. The main goal is to reduce or eliminate the pathological accumulations of Thd and dUrd to prevent further damage to mtDNA, which would cause a clinical stabilization or improvement [2,59]. It is expected that mtDNA depletion can be reverted upon restoration of nucleoside homeostasis, but other alterations like deletions and mutations are less likely to be reversible, so it is important to initiate the treatment as soon as possible [62]. Fortunately, therapeutic strategies do not need to be directed to the affected organs because Thd and dUrd have a systemic distribution [53].

One of the first therapeutic approaches that were attempted was the removal of excess amounts of Thd and dUrd form circulation by haemodialysis. In one study a significant reduction was observed, but this effect was transient and levels increased 3 hours post-dialysis. These deoxyribonucleosides have a high metabolic rate that exceeds the clearance by dialysis [29]. When the dialysis was repeated over time, a progressive reduction was observed, but there was no improvement on neurological symptoms and the disease progression was not stopped [63,64]. Another alternative was continuous ambulatory peritoneal dialysis (CAPD). Patients showed an improvement in gastrointestinal symptoms, but other major clinical features did not change [60]. An additional case report showed an improvement of gastrointestinal and neurological symptoms after CAPD, but the effects lasted only 15 months [65]. So although these approaches have some benefits, dialysis is a burdensome procedure and the therapeutic effects are short-term.

Another therapeutic strategy proposed was platelet transfusion, since those cells are rich in TP. This procedure elicited efficient recovery of functional TP enzyme and correction in Thd and dUrd levels. However, due to the short lifespan of platelets (a few days), these improvements were temporary and multiple treatment sessions were needed. Moreover, long-term platelet transfusion carries the risk of developing immune reactions and transmission of viral infections [63,66,67].

Allogenic hematopoietic stem cell transplantation (AHSCT) is an attractive therapeutic option because it offers the permanent correction of TP deficiency. Studies show that surviving patients had a reduction of plasma Thd and dUrd and presented clinical improvements, although residual disease manifestations persisted [68,69,70,71]. However, a compatible donor is not always available, and this procedure has an elevated mortality risk (62,5%) due to disease progression and post-operation complications such as graft versus host disease (GVHD) [70,71]. Moreover, a follow-up study suggested that the effect of AHSCT might be transient [72]. When patients are diagnosed they are usually in a poor clinical condition, so they have an impaired capacity to tolerate the transplant procedure, the aggressive conditioning and the immunosuppressive chemotherapy. Consequently, it is recommended that recruitment is restricted to patients in a stable clinical condition without irreversible end-stage disease that have an optimal donor [70,73].

Since the liver has a high TP expression, an orthotopic liver transplant (OLT) could be a useful therapy [74]. A case study showed that OLT was able to normalize Thd and dUrd levels and provide a mild improvement of the neurological symptoms, but damage in post-mitotic tissues was irreversible [75,76]. Like AHSCT, this procedure needs to be performed during the initial stages of the disease to achieve significant improvements. Moreover, this approach requires a suitable donor, long-term immunosuppression and involves transplantation related risks [77].

Enzyme replacement therapies (ERTs) have also been tested. To prolong the circulatory half-life of the enzyme and prevent immunogenic reactions, TP needs to be encapsulated before administration. A research group explored the use of polymeric enzyme-loaded nanoparticles. The *in vitro* results were promising, but further investigation is needed to confirm the feasibility of this approach [78]. The use of autologous erythrocyte-encapsulated TP (EETP) is under investigation and has Orphan Drug Designation by USA Food and Drug Administration (FDA) and European Medicines Agency (EMA) [79]. Autologous erythrocytes

are isolated from patients and loaded with recombinant TP via hypo-osmotic dialysis. Since Thd and dUrd can diffuse though the erythrocyte membrane, the excess of those nucleosides could be cleared by TP that is encapsulated inside them [79,80,81]. Preclinical studies showed that TP activity was maintained throughout the erythrocyte lifespan and that encapsulation prevented an immune response against TP [79]. Some patients received EETP under a compassionate use programme and showed clinical and metabolic improvements during years [81]. However, although Thd and dUrd levels were decreased, they were still detectable in plasma, so the reduction was suboptimal [79,82]. A clinical trial to evaluate the efficacy of this therapy has recently started [83]. With EETP patients need periodical infusions, so it is not a permanent cure and it is expensive [79]. Nevertheless, it can be useful to slow MNGIE progression and stabilize the clinical condition, which increases the patients chance of being eligible for AHSCT or OLT [53].

There are other potential strategies that have been proposed, such as the pharmacological inhibition of Thd and dUrd renal reabsorption. This would increase urinary clearance, thus decreasing circulating levels. The main challenge is to find a drug that selectively inhibits the pyrimidine deoxyribonucleoside reabsorption without hindering the physiological renal transport of other compounds [62]. Another potential approach is the increase of dCtd levels in order to prevent mitochondrial dCTP depletion. Preclinical studies showed that the administration of dCtd, or an inhibitor of its catabolism, could prevent mitochondrial dCTP depletion and restore mtDNA copy number [43,84]. However, this strategy has no effect in excess Thd and dUrd levels. Moreover, the optimal delivery method has to be considered. Hence, this approach is proposed as a maintenance treatment before a more suitable alternative can be applied [62].

2.1.9. MNGIE mouse model

The existence of animal models is crucial for the development of disease therapies. However, it is important to consider the biological differences between humans and animal models. In mice, the metabolism of Thd and dUrd can be performed by three different enzymes: TP, uridine phosphorylase 1 (UPP1), and uridine phosphorylase 2 (UPP2) [85].

Taking this into account, a scientific group established a murine model based on a double knock-out (dKO) *Tymp*$^{-/-}$ *Upp1*$^{-/-}$. Although these animals presented some MNGIE features, they were not ideal. The increase of Thd and dUrd levels in plasma was less pronounced compared to humans, and they had no detectable mtDNA abnormalities in brain and muscle tissues. The lack of mtDNA depletion in mice may be the result of different repair and replication pathways that are not affected by an increase of Thd and dUrd concentrations inside mitochondria [86].

Later, another group created a second dKO *Tymp*$^{-/-}$ *Upp1*$^{-/-}$ model to characterize the biochemical, genetic and histological features of MNGIE in mice. They initially generated *Tymp*$^{-/-}$ mice, which did not manifest a clinical phenotype and only showed minor alterations of nucleosides in liver and spleen [42]. Then those animals were crossed with *Upp1*$^{-/-}$ mice [87] to obtain a dKO. These animals displayed undetectable TP activity in all

tissues except the liver, which had a residual activity (17%) attributed to UPP2 expression. They had an increase of Thd and dUrd levels by 4-65-fold in all tissues, dNTP pool imbalances and partial mtDNA depletions. These mice developed kyphosis at the age of 5 months and showed abnormalities in the brain, but they had no gastrointestinal or skeletal muscle involvement [42,88]. It was hypothesized that this difference of clinical features between humans and mice was caused by various factors. First, the mice may not live long enough to accumulate sufficient mtDNA damage in most tissues (shorter lifespan). Second, the deoxyribonucleoside imbalance in less dramatic in mice than in humans so the impact on mtDNA is lower [42,44]. Lastly, mouse neuronal cells have a high TK2 expression, and this contributes to an increased dTTP production that accelerates mtDNA damage in nervous tissues [89]. The fact that this model presented mtDNA aberrations and the previous did not could be explained by limitations in the analytical method used in the first study [42].

To enhance the phenotype of this animal model, another group supplemented the diet of mutant mice with exogenous Thd and dUrd. This prolonged supplementation resulted in the acquisition of biochemical abnormalities and mtDNA depletion in the small intestine and brain. The animals presented gastrointestinal pathology, weight loss, muscle weakness, leukoencephalopathy and decreased survival [44].

2.2. Gene therapy

Gene therapy is an experimental technique that uses genetic materials to treat diseases. This can be addressed in several ways depending on the disorder, like introducing a functional copy of a gene, repairing a mutation, or inactivating a gene. The genetic material is often carried by vectors, which act as protective shields and allow the efficient delivery to the cells. There are two fundamental strategies for gene delivery: *ex vivo* and *in vivo*. *Ex vivo* gene therapy requires the harvest of autologous cells from a subject, their genetic modification in culture, and their reinfusion into the patient. On the other hand, *in vivo* gene therapy consists in the direct injection of the genetic material into the patient, so the genetic modification takes place inside the subject [90].

Since the introduction of gene therapy concept in the 1970s, the field has advanced considerably with notable clinical successes [90], despite significant and highly publicized setbacks early on [91,92]. The first trial based on genetic transfer in humans was done in 1990. They transduced tumour-infiltrating lymphocytes with γ-retrovirus vectors that carried the neomycin resistance gene, and infused them to patients with metastatic melanoma [93]. Shortly thereafter, the first gene therapy clinical trial was started. Two patients with adenosine deaminase severe combined immunodeficiency (ADA-SCID) where infused with T cells transduced with a retroviral vector that carried the *ADA* gene [94]. Within these 30 years, new vectors have been developed, and other modes of gene regulation and modification have been discovered [95]. There have been more than 3000 gene therapy clinical trials approved worldwide [96] (Fig. 4a) and 20 licensed products [97]. Cancer is by far the most common disease treated by gene therapy, followed by monogenic diseases [98] (Fig. 4b).

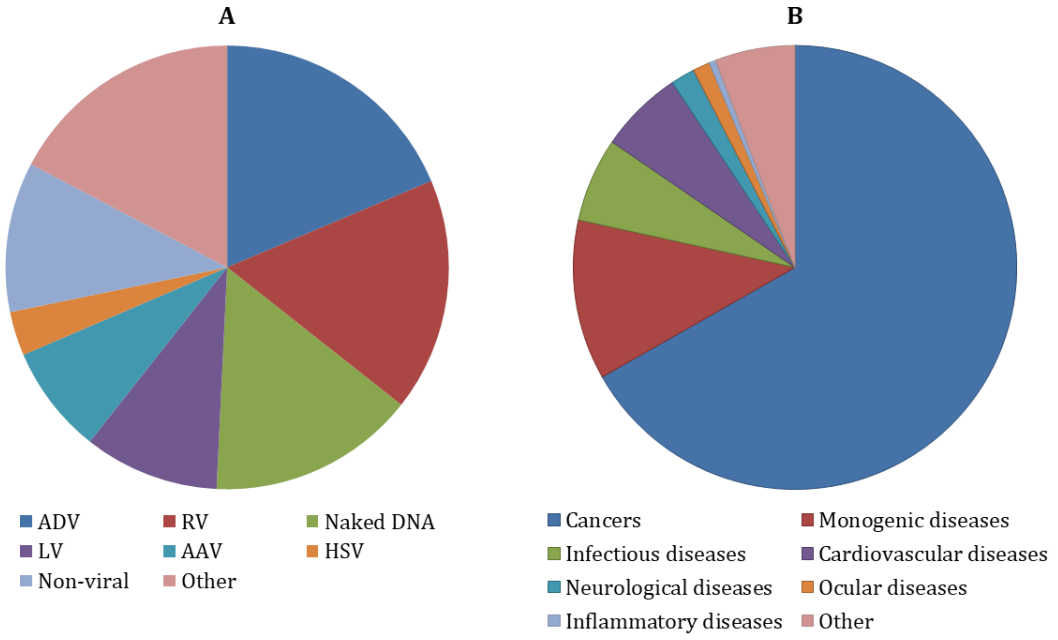

Figure 4. Statistics of gene therapy clinical trials. Classification of gene therapy clinical trials by (A) type of vectors [96] or (B) target diseases [98]. ADV: adenovirus; RV: γ-retrovirus; LV: lentivirus; AAV: adeno-associated virus; HSV: herpes simplex virus.

2.2.1. Gene therapy vectors

There are studies where the genetic material is delivered without a vector. But naked DNA/RNA administered intravenously is rapidly degraded by blood serum nucleases, so it is a very inefficient process [99]. The use of hydrodynamic injection increases the chance of cell entry of the naked genetic material. The rapid infusion of a large volume of liquid elevates the circulatory hydrostatic pressure and forces the permeabilization of the cellular membranes. Another strategy is the use of electroporation, which consists in the application of pulsating electric fields after the injection of the genetic material to permeabilize the cellular membranes. This procedure has been tested in many organs, but the most common are the skeletal muscle and tumours [100,101].

However, most gene therapy strategies use vectors to protect and deliver efficiently the genetic material. The most used vectors are viruses, because they have the natural capacity to carry nucleic acids and deliver them to cells at high efficiency rates. But nowadays there are many platforms that are based on non-viral vectors [100].

2.2.1.1. Viral vectors

2.2.1.1.1. Retroviral vectors

Retroviruses (RVs) are relatively big viral particles (80-120 nm) that have 2 single-stranded RNA (ssRNA) chains as a genome. Their distinctive feature is the presence of a proteolipidic envelope that encapsulates the viral capsid. The glycoproteins of the envelope attach to the host cell receptors and the viral particles are released into the cytoplasm. Then, they disassemble and start the reverse transcription of the ssRNA to produce double-stranded DNA (dsDNA), which enters the nucleus and integrates into the host genome. Once inserted, the viral genome can be transcribed and translated to generate the regulatory and structural proteins needed to produce new viral particles [100,102]. Although each type of RV has different integration site preferences, the integration patterns are random and dependant on the cell type [103].

Gamma-retrovirus (γ-RV)

The first vector used for gene therapy was the γ-RV. Its genome contains three genes, which encode for the capsid proteins (*gag*), the envelope proteins (*env*), and the reverse transcriptase, protease and integrase (*pol*). They are flanked by long terminal repeats (LTRs), which function as enhancers and promoters [100,102].

Its ability to transfer genes was discovered as a natural mechanism by which these viruses cause tumours in vertebrates [104,105]. This led to the construction of γ-RVs based on Moloney murine leukaemia virus (MMLV), which were useful for the *ex vivo* transduction of a wide range of mammalian cells [106,107]. MMLV integrates in a non-random pattern that favours regulatory elements (promoters, enhancers, conserved non-coding regions) [103]. The first generation vectors substituted the viral genes by the transgene cassette, but the

LTRs were maintained. These vectors were used in a clinical trial for X-linked SCID (X-SCID), but some patients developed leukaemia due to insertional mutagenesis. The viral genome was inserted near the *LMO2* oncogene and the LTRs promoted its over-expression [92]. To solve this, self-inactivating (SIN) vectors were developed. They have truncated inactive LTRs to decrease the risk of activating a nearby gene [108] (Table 4).

Since they were the first vectors developed, there are several γ-RV-mediated therapies in the market right now. The Philippine FDA authorized in 2007 the marketing of Rexin-G™, and USA FDA approved it in 2010. This product is based on a γ-RV vector that contains the cytocidal cyclin G1 construct, and is prescribed to patients with metastatic pancreatic cancer [109]. In 2016 EMA approved Strimvelis™, an *ex vivo* gene therapy for ADA-SCID. Autologous hematopoietic stem cells (AHSCs) are transduced with a γ-RV that carries the *ADA* gene [110]. During the same year EMA also approved Zalmoxis™, an *ex vivo* gene therapy for preventing GVHD due to HSCT with a partially matched donor. Allogeneic donor T cells are transduced with a γ-RV that carries an inducible suicidal gene. If severe GVHD develops, T cells can be killed using a drug and stop the immune rejection [111]. Next, in 2017 USA FDA approved Yescarta™, an *ex vivo* γ-RV-based therapy for B cell lymphoma. Autologous chimeric antigen-receptor T (CAR-T) cells are engineered to target CD19-positive cells [112,113]. Lastly, in 2017 the South Korea Ministry of Food and Drug Safety (MFDS) approved Invossa™, an *ex vivo* γ-RV therapy for knee osteoarthritis. Allogenic chondrocytes are engineered to overexpress transforming growth factor β1 [114].

Lentivirus (LV)

LVs are a family of complex RVs. Its genome has the same structure as γ-RVs, but they have additional genes that encode proteins important for the viral replication, binding, infection and release. Other RVs can only enter the nucleus when the nuclear membrane is dissolved during cell division [115], but LVs can enter through nuclear pores, so they can also infect quiescent cells [102,116].

The majority of LV vectors are based on the human immunodeficiency virus (HIV), but they can be derived from others like the feline immunodeficiency virus (FIV) or equine infectious anaemia virus (EIAV) [116]. HIVs prefer to integrate within coding regions and avoid regulatory regions, so they are less likely to cause the over-expression of proto-oncogenes [103]. Even so, SIN LVs were also developed to inactivate LTRs and improve their safety profile (Table 4).

In 2017 USA FDA approved Kymriah™, an *ex vivo* LV therapy for acute lymphoblastic leukaemia. Autologous CAR-T cells are engineered to target CD19-positive cells [113]. Also, in 2019 Zynteglo™ got a conditional approval by EMA. It is an *ex vivo* gene therapy were AHSCs are transduced with a LV vector that carries the *HBB* gene. It is prescribed to patients with β-thalassemia [97].

2.2.1.1.2. Adenoviral vectors

Adenoviruses (ADVs) are relatively big viral particles (70-100 nm) that have a dsDNA chain as a genome. They have two major transcription regions that encode proteins essential for DNA replication and viral packaging. These genes are flanked by two inverted terminal repeats (ITRs) [90]. Upon infection, the viral genome remains episomal in host cells, so these vectors are well suited for slow dividing or quiescent cells [117].

The first generation of ADV vectors retained some viral genes in the genome. They had a robust transduction in many tissues, but only a transient expression because these vectors were very immunogenic [118,119,120]. With time, these first generation vectors were improved by deleting more viral genes. In 1999, during a clinical trial for ornithine transcarbamylase deficiency, a subject treated with a high dose of ADV vector died due to a severe systemic inflammatory response and organ failure. This supposed a big drawback in the gene therapy field at that time [91,121]. More recently, ADV vectors without viral genes were developed (gutted ADV vectors). They provided a long-term transgene expression with reduced immunogenicity [122]. Another approach is to take advantage of their high immunogenicity and use them as oncolytic viruses or vaccine carriers [123,124,125] (Table 4).

The China Food and Drug Administration (CFDA) authorized in 2003 the marketing of Gendicine™. This product is based on a ADV vector that contains the *p53* gene, and is prescribed to patients with head and neck squamous cell carcinoma [126]. Oncorine™ was also approved by CFDA in 2005. This product is based on an oncolytic ADV prescribed to patients with late-stage cancer [127].

2.2.1.1.3. Herpes Simplex Virus (HSV) Vectors

HSVs are large viral particles (150-200 nm) that have a dsDNA chain as a genome. They contain approximately 90 genes that are flanked by inverted repeated sequences. HSVs have an envelope that confers a broad host cell range, although it has a predilection for epithelial and neuronal cells. They are highly infectious and can switch between latent and lytic modes [128].

Replication-defective HSVs can establish a long-term latency in neurons, so they have been used to treat neurological disorders. On the other hand, replication-competent HSVs have been developed as oncolytic viruses for cancer treatment. They have been used in several trials for malignant brain tumours [129,130,131,132] (Table 4).

The USA FDA authorized in 2015 the marketing of Imlygic™, which was subsequently approved in Europe and Australia in 2016. This product is based on an oncolytic HSV-1 that contains the granulocyte-macrophage colony-stimulating factor gene. It is used to treat solid tumours like melanomas [133].

2.2.1.1.4. Adeno-associated virus (AAV) vectors

AAVs are small (20-25 nm) non-pathogenic parvoviruses that naturally infect humans and other mammalians [100,117,134]. They were first identified as contaminants in ADV vector preparations [135], but it was not until 20 years later that the AAV2 viral genome was cloned and sequenced [136,137] (Table 4).

Vector	γ-RV	LV	ADV	HSV	AAV
Genome	ssRNA	ssRNA	dsDNA	dsDNA	ssDNA
Diameter	80-120 nm	80-120 nm	70-100 nm	150-200 nm	20-25 nm
Insert size limit	7-10 kb	10 kb	30 kb	40-50 kb	4,7 kb
Host cells	Non-dividing	Dividing and non-dividing	Dividing and non-dividing	Dividing and non-dividing	Dividing and non-dividing
Viral genome	Integrative	Integrative	Episomal	Episomal	Episomal (>90%)
Expression	Stable	Stable	Transient	Transient	Transient
Advantages	Not immunogenic	Not immunogenic High transgene expression	High transgene expression High titer Broad tropism	Broad tropism Strong neurotropism Lytic/latent modes	Non-pathogenic Less immunogenic Low risk of genotoxicity Long-term expression in quiescent cells Broad tropism
Limitations	Low transgene expression Low titer Risk of genotoxicity	Risk of genotoxicity	Very immunogenic Lethal at high doses	Very immunogenic Challenging manufacturing process	Pre-existing immunity

Table 4. Comparison of the properties of the main viral vectors [90]

2.2.1.1.4.1. AAV structure

Genome

AAVs have a single-stranded DNA (ssDNA) genome of approximately 4,7 kb. It contains three genes (*rep*, *cap*, *aap*) and three promoters (p5, p19, p40). Multiple proteins can be synthesized through the different promoters, differential splicing, and alternative translation start sites. These genes are flanked by two ITRs [137,138,139] (Fig. 5).

- **ITRs**: 145 nucleotide-long sequences required for packaging, replication and chromosomal integration. The first 125 nucleotides are palindromic and form a T-shaped hairpin structure that is used as a self-prime to synthesize the second strand during replication [137,140,141]. In the presence of Rep proteins, ITRs have enhancer activity [142].
- ***rep* gene**: Encodes four proteins (Rep78, Rep68, Rep52, Rep40) involved in the viral genome replication, transcriptional regulation and packaging [100,117,137,138,139,142].
- ***cap* gene**: Encodes three proteins (Vp1, Vp2, and Vp3) that constitute the viral capsid. They mediate the host cell recognition and internalization [100,137,138,139].
- ***aap* gene**: Encodes the assembly-activating protein (Aap), which promotes the assembly and maturation of the capsid and facilitates its nuclear import [143,144,145]. However, this protein is not essential in all AAV serotypes [146]. This gene is located in an alternate reading frame overlapping the *cap* gene [143].

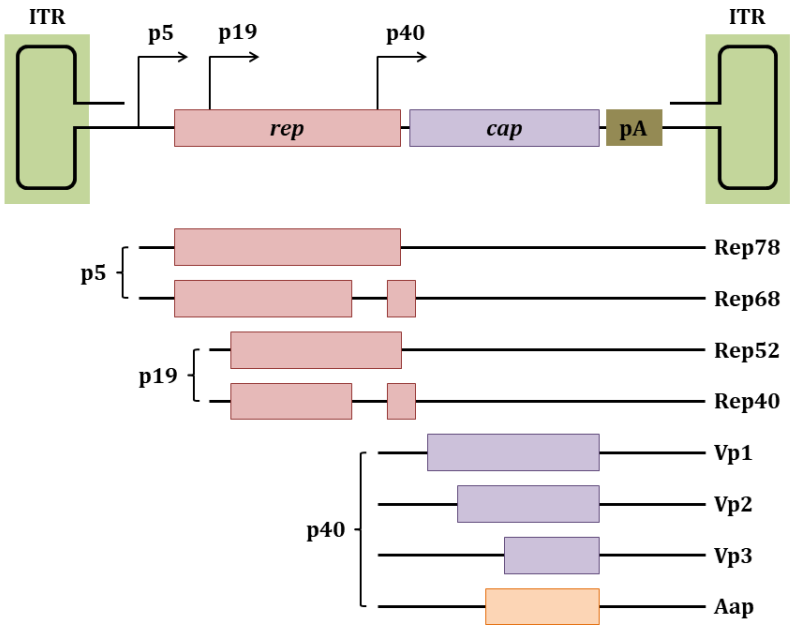

Figure 5. AAV genome structure. p: promoter.

Capsid

AAVs have an icosahedral viral capsid composed by 60 Vp proteins. The three types of Vp proteins are present in this structure in a molar ratio of 1:1:10 (Vp1:Vp2:Vp3) [100,117,147]. Vp3 is the most abundant (80%) and is responsible of the binding to the virus to different receptors and essential for the formation of new virus particles. The natural variability of this protein accounts for the differential viral tropism [148]. Vp1 is essential for AAV infection [149], and Vp2 is the major responsible for nuclear transfer [150].

2.2.1.1.4.2. AAV viral cycle

A) Host cell recognition and infection

The viral capsid proteins interact with carbohydrates and receptors present in the surface of the target cells. Each AAV serotype has preference for different cell surface receptors (such as sialic acid, galactose, or heparan sulfate), which explains their distinct tropism [117,151,152]. But there is a universal primary receptor used by most AAV serotypes called AAVR [153]. The binding of AAV to its cell surface receptors triggers intracellular signalling pathways that stimulate AAV internalization [152] (Fig. 6).

B) Endocytosis

The first described internalization mechanism for AAV was clathrin-mediated endocytosis [154,155], but other endocytic mechanisms were described later [156,157,158]. Certain uptake pathways lead to a successful transduction, while others end with AAV degradation before reaching the nucleus [156,158]. The endosomal intracellular trafficking pathways are different depending of the cell type and AAV serotype. Once the viral particles are inside endosomes, they travel to the Golgi apparatus [156,159]. During the late endosomal stages, the capsid proteins Vp1 and Vp2 undergo a conformational change that leads to the exposure of their N-terminal ends, which contain nuclear localization signals (NLS). Vp1 N-terminus also has a domain of phospholipase A2 (PLA2), an enzyme that disrupts the endosomal membrane and facilitates the release of the viral particles to the cytosol [160,161] (Fig. 6).

C) Nuclear entry

The intact viral particles enter the nucleus through the nuclear pore complex (NPC) [155,162,163]. The nuclear AAV transport is a slow and inefficient process, as only 1-2% of the internalized AAVs actually enter the nucleus and it takes 2-13 hours. Most viral particles are eventually degraded by host cell proteasomes and presented as antigens via the major histocompatibility complex (MHC) class I pathway [164]. For this reason, nuclear entry is a major rate-limiting step in the AAV infection pathway [165]. Once AAVs are inside the nucleus the capsid disassembles and the ssDNA genome is released. Then, the cellular DNA polymerases slowly convert the viral genome to a dsDNA using ITRs as primers. This is another rate-limiting step in the AAV infection in the absence of a helper virus [166]. Alternatively, if the cell was co-infected with 2 AAVs that contained complementary ssDNAs, the two strands can be annealed and form a dsDNA directly [167]. These dsDNA molecules are then circularized and concatemerized due to homologous recombination

(HR) of ITRs, and remain as episomes. But the viral genome can also be integrated into a specific region on human chromosome 19 (19q13.42), called AAVS1 [168,169] (Fig. 6).

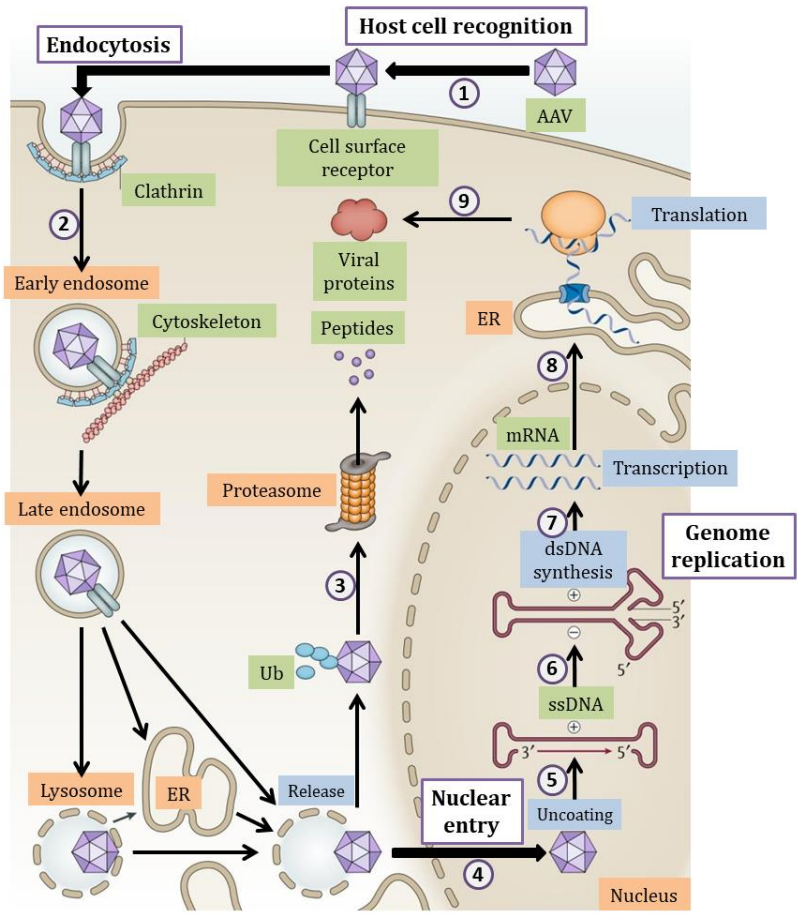

Figure 6. **AAV viral cycle.** AAVs bind to receptors and co-receptors on the surface of the target cells (1) and are internalized through endocytosis (2). Due to conformational changes, AAVs are released into the cytosol. From then on, they can either be ubiquitinated and targeted for proteasome-mediated degradation (3), or introduced into the nucleus (4). Once inside the nucleus, AAV particles are uncoated and the viral genome is released (5). Then the second strand synthesis starts and the genome replication is initiated (6). Next, the genomes are transcribed (7) and the mRNAs are transported to the cytosol (8). Finally, the mRNAs are translated and the viral proteins are generated (9). Figure adapted from [170]. ER: endoplasmic reticulum; Ub: ubiquitin.

D) Viral genome replication

AAVs cannot establish a productive viral infection alone; they need the replication machinery of helper viruses such as ADVs or HSVs. In presence of a helper virus, the transcription of the AAV promoters is up-regulated and the AAV proteins are produced. In absence of a helper virus, AAV viral promoters are practically inactive and AAV remains latent [171]. Once the Rep proteins are expressed and imported to the nucleus, AAV DNA is replicated by the rolling hairpin mechanism [172,173]. Since ITRs are identical, this process occurs equally well from both ends, so both the sense and antisense DNA strands are synthesized [139,140] (Fig. 6,7).

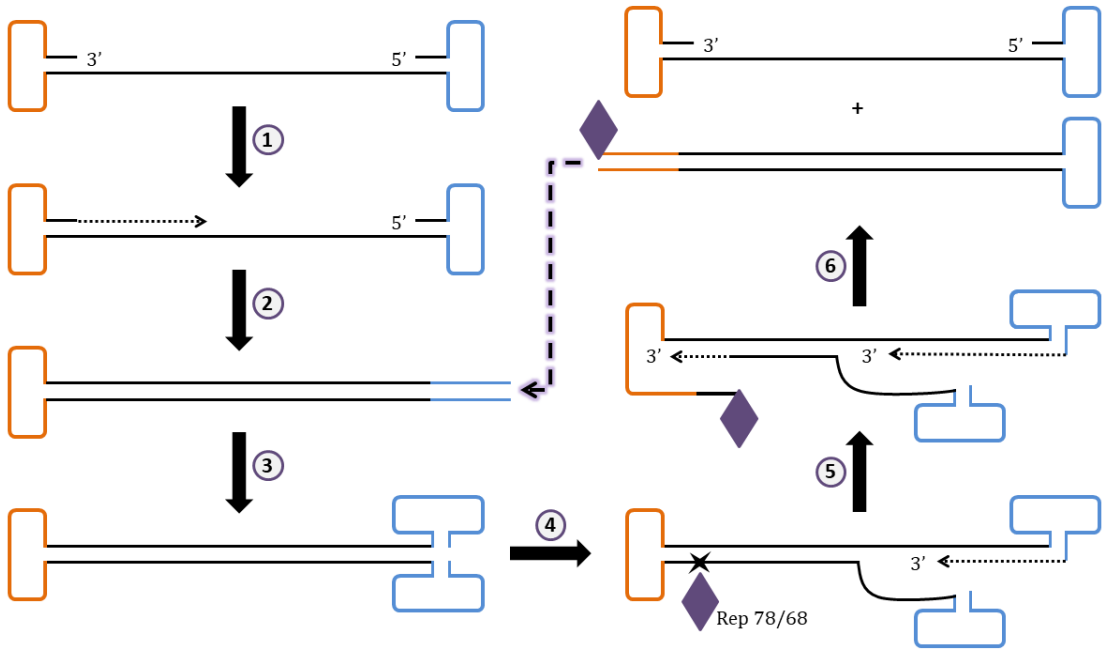

Figure 7. AAV genome replication by rolling hairpin mechanism. When the viral genome is released inside the nucleus, host DNA polymerases start the second strand synthesis (1), until a duplex molecule with a terminal hairpin is obtained (2). Then, the non-hairpined extreme isomerizes and forms a double hairpin (3). Next, Rep78/68 proteins bind to the extreme with the single hairpin and produce a nick in the dsDNA (4). The cellular replication proteins start synthesizing DNA chains from the free 3'-ends (5). At the end, two molecules are obtained: an ssDNA and a dsDNA (6). The ssDNA is ready to be packaged in newly formed viral capsids, while the dsDNA is used as a template for the next replication round [172,173].

E) Generation of new viral particles

Vp and AAP proteins are synthesized and imported to the nucleus. There, AAP promote the empty capsid formation by assembling the three types of Vp [143,144,145]. Rep78/68 proteins attach themselves to newly formed ssDNAs and act as packaging signals. On the other hand, Rep52/40 mediate the entry of viral ssDNA genomes inside capsids. Both the sense and antisense DNA strands are packaged. Finally, the accumulated viral particles are liberated and induce the lysis of the host cell [174,175,176].

2.2.1.1.4.3. AAV vectors characteristics

As a result of its simple structure, unique biology, and no known pathogenicity, AAV vectors could become the best choice for many gene therapy applications. The design of AAV vectors should always aim to maximize the potency of transgene expression in order to decrease the dose and reduce the risk of dose-related immune responses and toxicity [177]. This is also important because it is known that most AAV vectors are less effective in humans than in mice [178].

Recombinant AAVs (rAAVs) are obtained when the viral genes are substituted by the genetic material of interest. The only part of the viral genome that is maintained are the ITRs, as they are essential for the genome packaging [117,179]. The AAV2 serotype was the first one used as a vector. The wild type AAV2 genome was cloned into a plasmid backbone and the viral genes were replaced with a marker gene and a promoter [141].

The rAAVs used nowadays are mostly generated by pseudotyping, which means they are generated by the combination of a genome and a capsid from different serotype origins. These recombinant vectors are referred as "rAAVx/y", where "x" is the number of the ITRs serotype, and "y" refers to the number of the capsid serotype. Since ITRs are highly conserved between serotypes and do not influence the viral tropism, the AAV2 ITRs are the most frequently used [117,180,181].

2.2.1.1.4.4. rAAV vector manufacturing

A) Production

The assembly of rAAV vectors in cultured cells requires the supply of three components: the AAV expression plasmid, which contains the therapeutic cassette, flanked by ITRs; the AAV viral genes; and the helper virus genes. There are commercial AAV expression plasmids that are ready to be cloned with the desired therapeutic cassette [182,183].

The most common method to produce rAAVs in a laboratory-scale is by co-transfection of HEK-293 cells. The AAV expression plasmid is introduced with one or two helper plasmids that contain the AAV and ADV genes. These cells are able to produce AAV capsids and package them with the therapeutic construct flanked by ITRs. This method is rapid, safe and efficient, but it is not adequate for the large-scale production needed for clinical applications [184].

One method to scale-up the production is to use a stable packaging cell line that already contains the AAV and ADV genes inside their genome, so the cells continuously express all required components for AAV packaging in sufficient quantities. The only necessary step would be the transfection of the desired AAV expression plasmid. Unfortunately, these cell lines turned out to be unstable due to the inherent toxicity of the viral genes. To solve this, some systems to control the expression of the viral genes were designed. Still, these cell lines remained unstable and yielded low AAV titers, so it is not the best method to upscale rAAV production [100].

Another option is the use of HSV strains that contain the AAV genes. The packaging cells are infected with this "super-helper virus" and get the necessary components for rAAV packaging. The production with this method is superior because infection is more efficient that transfection [185]. The packaging cells are also infected with a second HSV that carries the therapeutic cassette flanked by ITRs. With this double infection, the packaging cell lines are able to produce rAAV vectors at a larger scale [186].

Alternatively, SF9 insect cells can be infected with a combination of baculoviruses that contain the AAV genes and the therapeutic transgene flanked by ITRs. This system has yielded high amounts of rAAV vectors in bioreactors [187].

B) Purification

AAVs have a high chemical and physical stability, so they can be subjected to rigorous methods for purification and concentration. This is a major advantage for bioreactor scale production, which is required for clinical application [100]. Purification is a crucial step, not only to remove residual cells and viral contaminants, but also to enhance the infectious potency of rAAV yields [100].

The first method used was the caesium chloride (CsCl) gradient centrifugation, but this chemical component caused the loss of rAAV infectivity [100]. Nowadays, the method of choice is the isosmotic iodixanol density gradients. The packaging cells are lysed and the crude cell extracts are treated with benzonase to digest cell-derived and non-encapsulated viral nucleic acids. Then the extract is loaded on the iodixanol step gradient for subsequent ultracentrifugation. Finally, the rAAV particles are collected and further purified by chromatography and dialysis. They can also be concentrated by ultracentrifugation [182,188].

C) Quantification

The choice method to quantify rAAV vector yields is the determination of AAV genome copy (gc) numbers by quantitative PCR. But some of the rAAV vector particles can be defective or non-infectious, so the valuable information is the amount of infective rAAV particles. However, this is difficult to evaluate since it is dependent on the transgene used, the target cell line, and the transduction conditions. In general, transduction of cell lines in culture is extremely inefficient. [100].

2.2.1.1.4.5. Serotypes and tropism

In gene therapy it is often desirable to direct the transgene expression to a specific cell type or tissue. This targeting allows higher transduction efficiency at lower rAAV doses and diminishes the risk of adverse effects due to transduction of the incorrect tissue [179]. This can be attained at a certain level by selecting the appropriate AAV serotype, although there are cell types that are difficult to transduce with any AAV [151]. To date, a total of 12 natural AAV serotypes have been discovered from human and non-human primates (NHPs) (AAV1-AAV12). They are classified based on shared sequence homology and tropism [100,152,189,190,191] (Table 5). Serotype also influences cell trafficking efficiency, capsid uncoating and gene expression [139].

rAAV vectors can be customized to improve transduction efficacy, change immunological profiles and enhance a certain tropism [179,170]. Capsids can be engineered by either screening mutated capsid variants, or by coating capsids with certain molecules. It is an effective procedure, but it is laborious and requires an extensive knowledge of the AAV

capsid biology, structure and function. To date, there are more than 100 rAAV variants [117,181,170,192,193].

There are other methods to improve the tissue specificity of rAAV vectors. The systemic delivery of rAAV often results in the accumulation in the liver, which is not always desirable. If they are administered locally it is less likely that cells of unwanted tissues are transduced. Moreover, the use of cell-specific promoters can also help direct the transgene expression. Alternatively, the insertion of binding proteins into the capsids can confer tissue selectivity [151,170]. Additionally, it has been reported that codon usage contributes to tissue-specific expression, so codon optimization is a valuable tool [170,194].

Serotype	Species origin	Primary receptor	Co-receptor	Human tissue tropism
AAV1	Monkey	Sialic acid	AAVR	Muscle, central nervous system (CNS), heart
AAV2	Human	Heparin	Integrin, FGFR, HGFR, LamR, AAVR	Liver, CNS, muscle
AAV3	Human	Heparin	FGFR, HGFR, LamR, AAVR	Muscle, stem cells
AAV4	Monkey	Sialic acid	Unknown	Eye, CNS
AAV5	Human	Sialic acid	PDGFR, AAVR	CNS, lung, eye
AAV6	Human	Sialic acid, heparin	EGFR, AAVR	Muscle, CNS, heart, lung
AAV7	Monkey	Unknown	Unknown	Muscle, CNS
AAV8	Monkey	Unknown	LamR, AAVR	Liver, muscle, pancreas, CNS
AAV9	Human	Galactose	LamR, AAVR	Every tissue
AAV10	Monkey	Unknown	Unknown	Muscle
AAV11	Monkey	Unknown	Unknown	Unknown
AAV12	Human	Unknown	Unknown	Nasal

Table 5. Characteristics of the 12 natural AAVs known to date [170]

2.2.1.1.4.6. AAVs for liver gene therapy

The liver has many properties that make it an ideal target for AAV gene therapy. It plays an important role in metabolism regulation, detoxification and synthesis of plasma proteins. The majority of these functions are done by hepatocytes, which constitute approximately 70% of the liver. This organ has easy access to the circulatory system, so it could work as an efficient factory to systemically distribute secreted proteins. Moreover, liver presents a highly fenestrated endothelium, which favours the uptake of many molecules from circulation. Due to its anatomical features, rAAV vectors administered

systemically naturally accumulate in the liver, so transduction is much more efficient than in other organs. Additionally, hepatocytes have a slow turnover, so although AAV are rarely integrative, the vector genome could be maintained for a long time [181,195].

rAAV vectors are ideal for *in vivo* liver gene therapy because they can efficiently transduce hepatocytes and maintain the transgene expression for many years. Moreover, they are less immunogenic than other vectors [196] and they present a low risk of insertional mutagenesis [169,197,198]. Furthermore, studies have shown that hepatocyte-restricted transgene expression at high levels can induce transgene-specific immunological tolerance [199,200,201,202,203,204,205,206,207].

A biodistribution study revealed that AAV7, AAV8 and AAV9 have a natural affinity for murine hepatocytes [191,208]. However, in human hepatocytes other serotypes, like an engineered AAV3 (AAV3B), have shown better transduction efficiencies [209]. It is clear that liver tropism of a specific AAV serotype is not necessarily conserved across different species. For this reason, it is important to consider that the efficacy an rAAV tested in mice may be significantly lower in humans [151,181,178]. Several groups have worked on the development of rAAVs to target human liver. A novel chimeric capsid derived from AAV3, named LK-03, was reported to transduce human hepatocytes at higher efficiencies than AAV8 [210]. Controversially, another report claimed that LK-03 is not superior to either AAV8 or AAV3B in transducing human hepatocytes [211].

Of the natural AAV capsids, AAV8 is considered one of the best for murine hepatocyte transduction. Studies show that peripheral vein administration results in an equivalent transduction compared to portal vein or splenic injection. Due to easy access and low technical difficulty, intravenous delivery is the preferred route of administration [202, 212]. Hepatocyte transduction efficiency differs both between mouse strains [199] and between genders [208,213]. Therefore, it is important to empirically determine therapeutic vector doses for the specific strain and gender to be used [181]. The AAV2/8 vector has been successfully used in many preclinical liver-directed therapies for diseases like haemophilia A [214,215], haemophilia B [216], lysosomal storage disorders [217], glycogen storage disease [218,219], familial hypercholesterolemia [220], and atherosclerosis [221].

2.2.1.1.4.7. Limitations of AAV gene transfer

- **Dilution of vector genome copy number with cell turnover**

Since AAV genomes remain mostly episomal they are diluted when cells divide, and consequently the transgene expression decreases. The rate of transgene loss is dependent on the turnover rate of the transduced cells [151]. Therefore, these vectors are not suitable to target a proliferative tissue if the goal is to have a long-term expression. However, AAVs are well suited to target slow dividing or quiescent cells, like myocytes, neurons, hepatocytes and retinal cells [62]. On a positive note, even if AAV genome remains episomal, it can persist for a long time and continues to express its transgene without any epigenetic silencing [222].

Although hepatocytes have a slow turnover, they are not quiescent, so it is expected to observe a decrease of the transgene expression in the long term [169]. This may not be important in mice, because they have a short lifespan, but in humans it could be a problem. Due to the immune system, it is not possible to re-administer the same vector twice, so strategies to avoid or minimize this immune response should be developed [181].

- **Limited packaging capacity**

AAVs are small viruses, so the size of the DNA they can contain is also small. Their maximum packaging capacity is approximately 5 kb, including the ITRs. Exceeding that limit results is incomplete packaging and reduction of the viral production yields [223,224]. If the desired transgene is bigger than this, the only option is to split the DNA sequence between different AAVs. If the target protein is composed of separate subunits, it can be possible to develop one AAV for each subunit [225]. Another option is to split the genetic sequence in various overlapping fragments. They can be packaged in different AAVs, and once they are inside the cell, the sequences recombine in the cytoplasm and form the complete sequence [226]. Alternatively, in some cases it is possible to reduce the size of transgenes to better adapt to the size limit. Trans-splicing vectors that form concatamers inside host cells have also been tested [226].

- **Delayed onset of gene expression**

Before the transgene contained in the rAAV vector can be expressed, the ssDNA must be converted to dsDNA by the DNA polymerases of the host cells. However, quiescent or slow-dividing cells do not have active DNA replication machinery, so this step takes a long time. As a consequence, the onset of transgene expression is delayed from days up to weeks until a plateau is reached [100].

One way to circumvent this is with the use of self-complementary AAV (scAAV) vectors, which contain a dimeric ssDNA genome that can fold back to a dsDNA once inside the host cell nucleus. scAAVs show an onset of transgene expression within hours and reach their plateau level within days. However, in practice this strategy causes the reduction of the packaging capacity to half [227].

- **Risk of genotoxicity by insertional mutagenesis**

Although wild-type (WT) AAV genomes remains mostly episomal, there is evidence that low frequency integration events occur, preferentially during cell division in AAVS1 [168,169,228]. Since this integration specificity is mediated by Rep proteins, rAAV vectors are expected to lose this property. In fact, it was reported that the occasional integration events of rAAV vectors do not have a genomic site preference, so they are expected to have a high safety profile for clinical application [197]. A study to examine the rAAV integration in clinical trials determined that low-level integration into random sites was not associated with toxicity [198]. Despite this, random integration of vector genomes may lead to loss or gain of function mutations that may alter cell functionality and homeostasis [177].

Several reports claimed that AAV genomic integration could result in the induction of hepatocellular carcinoma (HCC) in mice [229,230], but other reports denied it [231,232]. Additionally, there are no reports of HCC in larger animals, which were monitored during

a longer time [233,234]. Later, other reports claimed that AAV2 integrations were present in human HCC [235,236], but there was no direct evidence that the integrations where the cause, as it was possible that they were acquired after cell transformation [237,238]. At present there have been no reports of HCC in any AAV clinical trial done so far. However, the number of patients and the time of follow-up have not been large enough to draw definitive conclusions.

- **Sex-specific patterns**

The transduction efficiency of AAV is influenced by gender in some tissues. In the liver case, transgene expression was found to be 5-fold higher in male mice. One study found that liver transduction of castrated male mice was dramatically lowered, and suggested that AAVs transduce hepatocytes via an androgen-dependent pathway [213]. However, another study observed that the livers transduced with AAV8 vectors contained equivalent levels of vector genome between genders, suggesting that the differential transgene expression was mediated by post-transduction events [208]. Altogether, the sex-specific patterns must be carefully studied before clinical application.

- **Risk of transmission**

There is a concern of whether rAAV vectors can be transferred to other individuals by horizontal or vertical transmission.

Horizontal transmission can happen by the shedding of rAAV vector after *in vivo* gene transfer. Since vectors can be found in body fluids (blood, urine, semen) for several weeks after systemic administration [239], it may be possible to transmit them to another individual. Preclinical studies have demonstrated that vector sheds in fluids from NHPs were no longer infectious after 72 hours [240].

Vertical transmission refers to the risk of transduction of germline cells, which means that the rAAV DNA could potentially be transferred to the progeny. This was addressed in animal models and it was determined that rAAVs were not transmitted through the germline in male gametes [241]. Later these findings were corroborated in humans [242].

- **Immune response to AAVs**

AAV vectors are described as weakly immunogenic compared to other viral vectors. However, they do trigger the innate and adaptive immune systems through the viral capsid proteins and the nucleic acid sequence delivered. The presence of empty AAV capsids and contaminants derived from the manufacturing process may also influence the immunogenicity of rAAV vectors [177]. These immune responses are the major obstacle to achieve a safe long-term gene expression [151,196,243].

The capsid of the rAAV vectors is a replica of the WT AAVs, so it is expected that the same or similar immune responses will be triggered. However, WT AAV natural infection consists in low number of particles that enter via the airways, while rAAV therapies consist in large number of particles that are administered by other routes. Moreover, rAAV vectors do not actively replicate. These differences could play a role in shaping the immune response to rAAV vectors [244].

- Innate Immune Response

The innate immune response is the first line of defence against viral infection. Compared to other viral vectors, AAVs only trigger a mild innate immune response but it is sufficient to hinder long-term transgene expression [177,196]. Viral particles have pathogen-associated molecular patterns (PAMPs) that are recognized by pattern recognition receptors (PRRs) during cell transduction [243,245]. One type of PRRs are Toll-like receptors (TLRs), some of which are located in the cell surface. TLR2 has been reported as the cell surface sensor for AAV capsids in human liver cells [246]. When the viral particle enters the cell through endocytosis the capsid undergoes a conformational change that can expose the viral genome to TLR9, which recognizes sequences rich in CpG dinucleotides [247,248]. The interaction with TLRs triggers a signalling cascade to activate nuclear factor-κB and/or interferon regulatory factors, which stimulate production of pro-inflammatory cytokines. Additionally, soluble PRRs in the cytoplasm can also be stimulated by internalized AAVs [245,247]. Once the viral genome is inside the nucleus, dsRNAs are produced as AAV genome-derived replication intermediates. This stimulates the intracellular dsRNA sensors, which promote the production of type I interferon [249]. The generation of all these pro-inflammatory cytokines facilitates the immune cell recruitment and activation [250].

Changes in the vector genome sequence may diminish the innate immune response. TLR9 is an essential factor for AAV-mediated innate immunity, and since it recognizes CpG dinucleotides, transgenes with these patterns are more immunogenic [247,248,251]. Also, codon usage has been shown to contribute in the innate immune response to foreign DNA, so codon optimization is important.

- Adaptive Immune Response

The adaptive immune response is slow during the first infection and can be controlled to some extent. But after this, a population of memory T and B cells that are specific for AAV antigens remain latent in the organism. If there is a second infection, those cells are reactivated and rapidly eliminate the transduced cells. This is why AAV re-administration is challenging [243,252]. Over 90% of the human population has been or will be exposed to WT AAVs at one point of their life, so the presence of adaptive immune responses during an AAV-mediated therapy is to be expected.

This adaptive immune response can be directed to the capsid proteins or the transgene product. Viral proteins compose the capsid, so they will always be recognized by the immune system. The reaction against the transgene may happen if it encodes for a protein that the patient recognizes as foreign (for example in case of null mutations).

B cells

The natural exposure to WT AAVs causes the expansion of B cells that produce antibodies against AAV capsids (Fig. 8). Among them, neutralizing antibodies (NABs) are the main concern in AAV-mediated gene therapy, as they have the ability to block AAV cell internalization. About 70% of the population has NABs against AAV1 and AAV2, which are the most prevalent WT AAVs. Other serotypes are less prevalent, so NABs against them are less frequent [253,254,255]. However, NABs may also cross-react between serotypes and other similar parvoviruses [253,255,256]. This humoral immunity usually starts in childhood by the

age of 2 years. Also, maternal anti-AAV NABs can be found in newborns, although they disappear a few months after birth. Taking all of this into account, the general human population is only completely naïve to anti-AAV NABs during a small window of time, between 7 and 11 months of age [257,243,253]. It is also important to note that if the transgene encodes an extracellular protein, there may be NABs that recognize it and effectively block its function [250].

The presence of NABs can efficiently block the transduction of rAAV vectors. This problem was first evidenced in a haemophilia B clinical trial, where a subject with a high titer of NABs did not produce any detectable circulating factor IX (FIX) post-treatment [239]. This revealed the importance of pre-screening the levels of NABs in subjects before undergoing an AAV therapy. From then on, high titer of NABs is an exclusion criterion for enrolment in clinical trials that require intravenous AAV administration, so this humoral response is a significant limitation for AAV gene therapy [177]. It must also be taken into account that NABs are a problem with AAV re-administration. After the AAV treatment all patients will develop NABs against the AAV vector used, so a subsequent administration will not have the same therapeutic effect [250].

Different strategies have been proposed to overcome this problem. A different AAV serotype could be used, although it may be less optimal for the target tissue [258]. Moreover, NABs cross-reactivity makes it unlikely that simply switching the serotype alone would suffice [259]. More drastic alterations to the capsid structure are needed to evade NAB recognition, so rAAV capsid engineering may be a good option. Several rAAV variants with reduced sensitivity to NABs have been identified [256,260,261,262]. It has also been tested the use of empty capsids as decoys and the increase of rAAV doses to achieve the saturation of NABs. But this strategy increased antigen load, so it contributed to the activation of T cell mediated immune responses [252]. Another option is to envelop the AAV particles with exosomes or artificial nanoparticles to avoid the contact of the capsid with NABs. This was a very promising strategy, but these shielded rAAVs entered cells by another mechanism and the capsid tropism was lost [263,264]. Alternatively, it may be possible to temporary remove antibodies prior to AAV administration by performing a plasmapheresis or flushing the tissue [265,266]. Alternatively, the use of immunosuppressive drugs prior and/or after the rAAV treatment could block the generation of NABs during the first rAAV administration, so the same rAAV could remain effective after subsequent injections [250].

T cells

Antigen-presenting cells (APCs) are able to internalize AAVs, digest them, and present fragments of the capsid and transgene proteins as antigens. These antigens are presented in their cell surface through MHC molecules. Almost all cell types are able to present antigens, but there are professional APCs, like macrophages and dendritic cells (DCs), that are specialized in presenting antigens to T cells [243,250]. There are two major T cell groups: CD4+ T helper cells and CD8+ T cytotoxic cells. CD4+ T helper cells interact with the antigens presented by class II MHC molecules in APCs and trigger the adaptive immune response against them by producing cytokines that activate and expand CD8+ T cytotoxic cells and B cells. Then CD8+ T cytotoxic cells recognize antigens presented by AAV transduced cells via class I MHC molecules and kill them [164,243,250,267] (Fig. 8).

The structural differences of AAV serotypes and the differential cell entry mechanisms influence the T cell response against the transgene and the capsid [267,268]. The type of transgene and cells transduced, the dose and the route of administration also have an impact in the level of T cell response. All these variables determine if the immune response reaches the threshold needed to stimulate a functional CD8+ T cell response against the transduced cells [199,206,250].

The first problem is the immune response against the AAV capsid. Although animal models are extremely valuable to assess AAV safety and efficacy, they failed to predict the T cell reactivity to the AAV capsid in humans [177]. In a clinical trial for haemophilia B the subjects with the highest rAAV dose developed a transient and asymptomatic elevation of aspartate transaminase (AST) and alanine transaminase (ALT) (liver damage indicators) associated with a loss of *FIX* transgene expression. This was caused by capsid-specific CD8+ T cells, which targeted transduced hepatocytes [239,259,269]. This is a big problem because cells that have acquired the transgene are killed by the immune system. In a subsequent clinical trial this immune response was stopped with steroids. This treatment normalized the levels of AST and ALT and accomplished a stable transgene expression [205,270]. This demonstrated that the loss of transgene expression was due to the activation of capsid-specific CD8+ T cytotoxic cells. Moreover, there is a positive correlation between AAV dose and the cytotoxic immune response.

Current studies focus on immunosuppressive drugs to stop the CD8+ T cell-mediated cytotoxicity until the capsid has been degraded and cleared, but this does not solve the problem. Several strategies are being investigated to block capsid antigen presentation. One option is the reduction of the capsid antigen load by decreasing the rAAV dose administered. If the vector is modified to enhance transgene expression, a therapeutic level could be achieved at lower vector doses. With less rAAV vectors circulating the level of the immune response would be lower, and possibly it would not reach the threshold necessary to efficiently activate CD8+ T cytotoxic cells [271]. However, lower AAV doses are more likely to be completely blocked by NABs, so doses need to be carefully determined [252]. Another strategy is the inhibition of the proteasomal processing of the AAV capsid in order to prevent its presentation through MHC. These inhibitors would also enhance rAAV transduction by increasing nuclear translocation. However, these drugs are nonspecific and could produce unwanted side effects, and also their effect may depend on the target organ and AAV serotype used [164,272,273]. Alternatively, it has been explored the use of engineered rAAV capsids with modified surface tyrosine residues, which are targets to phosphorylation and ubiquitination. These modifications prevent capsid processing and presentation as antigens, and consequently attain a reduction of capsid-specific CD8+ T cell cytotoxicity against the transduced cells [274].

The second problem is the T cell immune response against the transgene product. CD8+ T cytotoxic cells detect transduced cells and kill them, which causes the loss of transgene expression [250,252]. More often than not the goal of AAV therapy is a long-term transgene expression, so the punctual use of immunosuppressive drugs after rAAV administration may not be sufficient. The AAV tropism for APCs is a determining factor, because if they are efficiently transduced, the T cell immune response is empowered. Some serotypes like AAV1 can transduce efficiently DCs and this results in greater transgene immunogenicity. However, AAV8 poorly transduces APCs and this correlates with a decreased CD8+ T cell

response [268,275]. In fact, AAV8 hepatocyte transduction results in the induction of antigen-specific CD4+ T cell tolerance to the transgene product [199,200,202]. Even pre-existing antibodies against the transgene product are eradicated after AAV8 liver gene transfer in animal models [276]. Taking this into account, it may be possible to exploit the tolerogenic environment of some tissues, like the liver, to induce a transgene-specific tolerance. The immune system must be highly regulated in order to rapidly respond to pathogenic infection while also preventing excessive inflammatory responses. To this end, several regulatory mechanisms exist to moderate the immune response by promoting immunological tolerance to antigens. In the liver, the immune response is predisposed toward tolerance. The presentation of antigens by liver cells triggers the expression of a variety of inhibitory cytokines and co-inhibitor ligands that results in a relatively poor CD4+ T helper cells activation. This contributes to incomplete activation of CD8+ T cytotoxic cells by either clonal exhaustion or premature death. Kupffer cells, myeloid-derived suppressor cells and liver DCs promote the activation of regulatory T cells and suppress effector T cells (CD4 and CD8). To induce tolerance, the vector has to efficiently transduce hepatocytes while avoiding the transduction of professional APCs. The innate immune response should be minimal and the transgene expression has to be powerful enough. If these requirements are met, the immune system becomes capable of recognizing the transgene product as a self-protein and will not react to its presence [199,200,201,202,203,204,205,206,207,250,277].

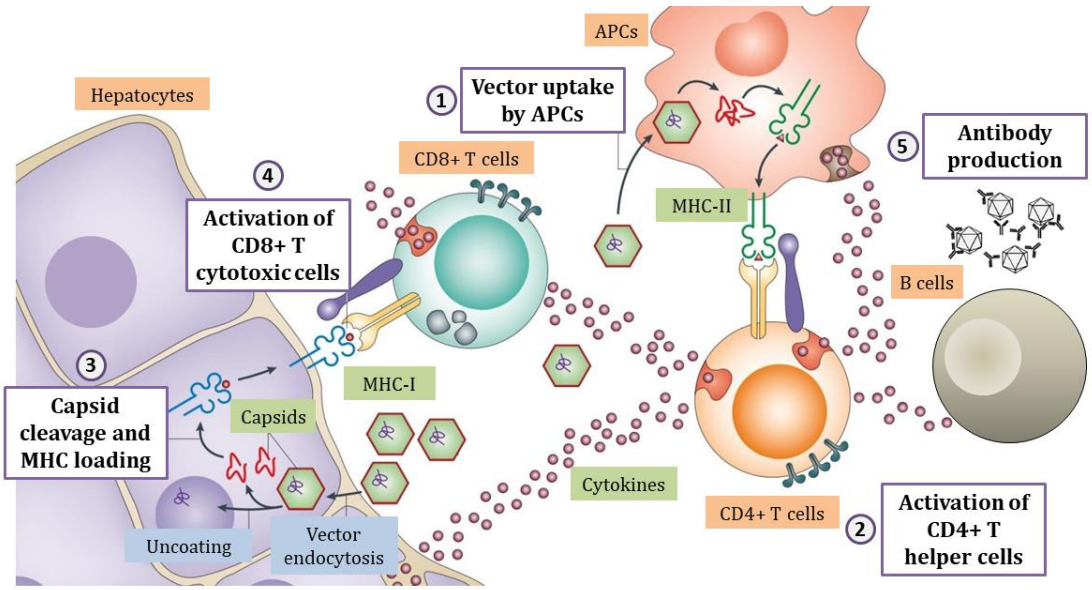

Figure 8. **Adaptive immune response towards AAVs.** When AAVs transduce cells the capsids are ubiquitinated and cleaved by the proteasome. The resulting peptides are loaded onto MHC and presented at the surface of the transduced cells. APCs present antigens by MHC class II (1). This presentation results in the activation of CD4+ T helper cells, which start producing cytokines. This triggers the activation of CD8+ T cytotoxic cells and B cells (2). On the other hand, the target cells (hepatocytes) present antigens by MHC class I (3). This presentation triggers their recognition by activated capsid-specific CD8+ T cytotoxic cells, which start killing the transduced cells (4). Moreover, B cells start producing and secreting anti-capsid antibodies (5). Figure adapted from [278].

2.2.1.1.4.8. Clinical trials with rAAV vectors

AAV vectors constitute a versatile viral vector platform that can be used in multiple diseases due to their unique biological and biophysical properties. Until now, rAAVs have proved to be safe and effective in many different clinical settings [205,270,271,279,280,281,282,283]. In 1995 the first clinical trial with rAAVs was initiated in cystic fibrosis patients [284]. From then on, more than 200 clinical trials with rAAV have been performed [96].

One of the best examples of promising liver gene transfer with rAAV vectors is haemophilia B. In a clinical trial, rAAV2 vectors carrying the human *FIX* transgene with a liver-specific promoter were used. The vectors were administered through the hepatic artery. This trial demonstrated for the first time that it was possible to transduce the human liver with rAAV vectors [239]. It also highlighted that AAV immunogenicity was an important limitation [259,269]. Gathering the previous results, a second clinical trial was started, this time using a scAAV2/8 vector carrying a codon-optimized *FIX* transgene. The vectors were administered intravenously through a peripheral vein. Moreover, a short course of immunosuppression was used to block potential immune responses against the vectors. In this case, the patients showed long-term transgene expression [205,270]. From then on, several clinical trials have started with different rAAV variants, and it is expected that in the near future a commercial product will be approved [285].

Although rAAV vectors are relatively new compared with the other viral vectors, there are already several AAV-mediated therapies in the market. EMA authorized in 2012 the marketing of Glybera™, the first-commercially approved gene therapy product for an inherited disorder in Europe. This product was based on a rAAV2/1 vector that contained the gene of lipoprotein lipase (LPL), and was indicated to a subset of patients with LPL deficiency [279,286]. However, since Glybera™ was not commercially profitable, it was retired from the market in 2017. Next, USA FDA authorized the marketing of Luxturna™ in 2017, the first-commercially approved gene therapy product for an inherited disorder in USA. It was subsequently approved by the EMA in 2018. This product is based on a rAAV2 vector that contains the *RPE65* gene, and is an orphan drug for Leber congenital amaurosis type 2 (LCA2) and retinitis pigmentosa type 20 (RP20) [280]. Lastly, the most recent gene therapy product authorized was Zolgensma™. USA FDA approved its commercialization in 2019, and is expected that EMA approves it in the near future. This product is based on a rAAV9 vector that contains the *SMN1* gene, and is prescribed to patients with spinal muscular atrophy type 1 [282].

2.2.1.2. Non-viral vectors

The use of non-viral vectors in gene therapy has increased over the past decade due to some advantages they present over viral vectors. First, they are generally less immunogenic, which greatly improves their tolerance and allows re-administration without immune reactions. Also, patients do not have pre-existing immunity against the synthetic vehicles. Second, they have an increased genetic capacity, which allows for multiple and large gene delivery and more complicated treatment designs. Finally, their mass-production is relatively simple, economic and reproducible. However, in return, non-viral vectors usually have a decreased delivery efficiency compared to viral vectors, as most synthetic vectors are unable to effectively trespass the multiple barriers of the organism. Moreover, due to their non-integrative nature, their effect is transient [101,287,288].

The first challenge for non-viral vectors is the potential degradation of the genetic material by endonucleases in physiological fluids and extracellular space. The vectors have to envelope the genetic material to protect it and improve their pharmacokinetics (increase its circulation time), while evading the immune system and renal clearance. At the same time, they must avoid aggregation and interaction with blood molecules. The second challenge is the ability to extravasate from the blood stream and reach the desired target tissue. Lastly, the vector has to mediate cell entry, endosomal escape, and nuclear entry. DNA has to be delivered into the nucleus to be expressed, while RNA can be delivered in the cytosol [101,288].

Non-viral vectors can be designed to carry different nucleic acids or proteins, but the delivery of messenger RNA (mRNA) has some particular advantages. First, mRNA does not need to be delivered inside the nucleus, which eliminates a potential limiting step. Second, mRNA has a better safety profile than DNA, because there is no risk of integration into the host genome. Third, the synthesis of the encoded protein is fast and its expression is transient. Finally, the production of *in vitro* transcribed mRNA can be easily standardized and up-scaled. They can be mass-produced at a low cost compared to other macromolecules. However, the use of mRNA for clinical purposes has been mainly limited by its physical instability, immunogenicity and difficult internalization due to its anionic nature. Various RNA chemical modifications have been developed to increase mRNA stability and efficiency. At present there are several mRNA-based clinical trials, most of them for cancer therapy [101,288,289].

Non-viral vectors encompass a wide range of compounds, like cationic lipids, polymers, dendrimers or peptides [287,288]. These compounds generally present a nanometric size, so they are commonly called nanoparticles (NPs). The interaction of NPs with cell membranes is determined by NP intrinsic properties (composition, size, surface charge, geometric conformation and concentration) and cell properties (cell type, healthy state and cell cycle phase). These factors will affect the NP internalization pathway, which ultimately determines their fate inside cells [290]. In the past decades, substantial advances have been achieved in the development of new delivery materials, as well as improved potency and stability of nucleic acids [288]. Multiple NPs have been developed all over the world, but in this work we will only focus on two compounds that exhibit promising results in RNA delivery: polymeric nanoparticles (PNPs) from *Grup d'Enginyeria de*

Materials (GEMAT) leaded by Dr. S. Borrós and located at *Institut Químic de Sarrià* (IQS) – *Universitat Ramon Llull* (URL) (*Barcelona, Spain*); and lipid nanoparticles (LNPs) from Acuitas Therapeutics® (*Vancouver, Canada*).

2.2.1.2.1. PNPs from GEMAT-IQS

GEMAT-IQS has developed a PNP formulation based on oligopeptide end-modified poly-(β-amino ester) (OM-pBAE) polymers for the delivery of nucleic acids. These polymers have the ability of condensing nucleic acids into discrete NPs, which facilitates cellular uptake and endosomal escape. pBAEs can be tailored for different applications by varying their structural components (backbone, side chain and terminal groups). This means that the surface charge of PNPs can be tuned, which permits the control of unwanted electrostatic interactions. These PNPs have a modular nature and can incorporate ligands by using polymerization or conjugation techniques. The OM-pBAEs used are designed to have more hydrophilic and cationic structures to encapsulate the genetic material by electrostatic interactions while maintaining the capacity of releasing the cargo inside cells. Once the PNPs are internalized, the polyplexes are decomplexed within 12 hours and the remaining polymers are recycled out of the cells. These PNPs have been used to deliver siRNA [291,292], mRNA [293,294], DNA [295,296], and oncolytic ADVs [297].

2.2.1.2.2. LNPs from Acuitas Therapeutics®

LNPs are currently the most popular and efficient non-viral delivery system. Cationic lipids are composed of a hydrophilic head and a hydrophobic tail. In aqueous environments, lipids spontaneously self-assemble into liposomes, enclosing the core of the hydrophobic tails and forming a spherical structure formed by lipid bilayers. Nucleic acids are molecules with a negative charge, so they are attracted to the cationic lipids and form electrostatic interactions. These lipid layers enhance cell internalization and protect nucleic acids from degradation and immunological responses [298,299]. However, LNPs with highly positive surface charges are prone to serum protein adsorption, aggregation, premature cargo release and clearance by immune cells [300,301]. Coating LNPs with hydrophilic molecules like poly(ethylene glycol) (PEG) has been shown to reduce immune stimulation and increase circulation time [302], although these LNPs may trigger immune responses upon re-administration [303].

Acuitas Therapeutics® has developed a LNP formulation that exhibits good results for *in vivo* RNA delivery to hepatocytes. These LNPs have a size of 70-100 nm and are composed of 4 different lipids: an ionizable amino lipid, a phospholipid, a cholesterol and a PEGylated lipid. At physiological pH these LNPs exhibit a net neutral surface charge. Once they are administered, the LNPs lose the PEGylated lipid from the surface and this enables the binding of apolipoprotein E (ApoE). This molecule facilitates the binding to hepatocytes and the LNP endocytosis [304,305]. During the endosomal maturation, the internal pH drops and the cationic lipids become positively charged. This results in the release of the nucleic acid payload to the cytoplasm. Finally, the lipids are rapidly cleared

as non-toxic metabolites. These LNPs were first developed for the delivery of siRNA [306,307,308], but later they were used for mRNA delivery [309,310,311,312]. Independently of the administration route, mRNA expression was mainly located in the liver. There was a strong protein expression in the first 24 hours, with the peak being at 4 hours. This effect was transient and disappeared within days, which demonstrates the rapid turnover of mRNA [309,312].

2.2.2. Gene therapy for MNGIE

The use of gene therapy is particularly promising for MNGIE for several reasons. First, MNIGE is a monogenic disease that can be treated with the correction of a sole gene. Second, there is no need to target a specific organ, as TP substrates can diffuse across plasma membranes. Lastly, an overall TP activity of 26-35% is sufficient to clear the systemic overload of Thd and dUrd, so it is very likely that the correction of a small cell population may be sufficient [62]. Moreover, gene therapy could circumvent many of the limitations of AHSCT, such as the difficulty in finding a suitable donor, or the toxicity of the conditioning procedures, the immunosuppressive drugs and GVHD.

Both LV and AAV vectors have been evaluated in preclinical studies performed at our institution by the group of Dr. R. Martí in collaboration with our group. The results show that gene therapy is a feasible and probably an optimal therapeutic option for MNGIE.

- **LV-mediated AHSC gene therapy**: This therapy consists in the isolation of AHSCs from MNGIE patients. Then the cells are transduced *ex vivo* by LV vectors carrying a functional copy of *TYMP*, and infused back into the patient [77]. Full or partial myeloablation could be required to allow the engraftment of the gene-corrected AHSCs, but immunosuppression would probably not be necessary since theoretically those cells are able to induce immune tolerance to TP [77]. As a proof of concept, B-lymphoblastoid cells from MNGIE patients were transduced with this LV vector, achieving the restoration of TP activity and normalizing Thd and dUrd homeostasis in the culture. This approach was also tested in the mice model [42]: Syngeneic bone marrow cells were obtained, transduced and transplanted into partially myeloblated mice. High levels of TP activity were observed in peripheral blood, which correlated with the reduction of Thd and dUrd plasma concentrations, even in animals with low percentage of transgene-expressing cells [88,313,314,315]. Results in the long-term follow up of 20 months were similar, which confirmed the lifelong biochemical correction [316]. However, as mentioned above, this procedure requires myelosuppression, which is risky in MNGIE patients, and this may limit the clinical applicability of this therapeutic strategy. Also, LV vectors can cause insertional mutagenesis, although a recent study reported that SIN LV vectors were safe [315].

- **AAV-mediated liver-directed gene therapy**: This therapy consists in the generation of rAAV2/8 vectors carrying the *TYMP* complementary DNA (cDNA). These vectors could be injected to the patients and would transduce hepatocytes *in vivo*. These rAAV2/8 vectors with the thyroxine-binding globulin (TBG) promoter

were used in the mice model [42] and achieved a long-term biochemical correction without adverse effects. dCTP depletion was also prevented. The liver TP activity was higher than the observed in control mice, but this did not have any apparent toxic effects [314]. However, a follow-up study determined that the biochemical correction was eventually lost in animals treated with the lower rAAV2/8 dose [317]. Next, it was found that the α-1 anti-trypsin (AAT) promoter led to higher TP activity in liver and more marked reductions in Thd and dUrd levels in blood and tissues. Moreover, these effects were detected at lower rAAV2/8 doses [318].

2.2.3. Genome editing

During the initial stages of gene therapy, the main goal was to deliver a functional gene to cells (gene addition). But the development of gene editing tools revolutionized the field. Nowadays there are many strategies that use these tools to repair mutations *in situ*, so the gene therapy field has broadened.

It is estimated that chromosomal double-strand breaks (DSBs) occur naturally in dividing cells at a frequency of ten events per cell per day. Their source can be endogenous (free radicals of oxidative metabolism, aberrant DNA metabolism) or exogenous (ionizing radiation, UV light). These DSBs cause genomic instability, which may result in cell death, carcinogenesis or accelerated aging [319,320,321,322]. For this reason, cells present two major pathways to repair DSBs: homology-directed repair (HDR) and non-homologous end-joining (NHEJ). Genome editing tools take advantage of these natural repair mechanisms to perform sequence-specific DNA editing in mammalian cells [95]. These tools are nucleases that induce a DSB in a specific genome locus, which is then repaired by either HDR or NHEJ [319,320,322].

The restoration of mutated genes by HR has always been an attractive goal. The first attempt was the introduction of a donor DNA molecule with the correct gene that had homology with the region of the genome that had to be repaired. It was expected that the presence of these homologies would trigger DNA recombination, and that the mutated gene would be substituted by the correct gene. However, this approach was heavily limited by the inefficiency of HR in mammalian cells and animal models [323]. But the discovery that DSBs could raise the incidence of HDR by several orders of magnitude revived this field. The first tools used to generate DSB to enhance HDR were "meganucleases", which are endonucleases with a large and extremely rare recognition site. However, these nucleases could only target a predetermined sequence, so they had a limited use [324,325,326,327]. Although they can be engineered, "meganucleases" have been overshadowed by artificial nucleases with better performance. But even with the new enhanced tools, HDR is mainly active in proliferating cells so this approach is not feasible in all tissues [319,322,328].

Genome editing approaches can overcome many of the drawbacks of the classical gene therapy based on gene addition. First, insertional mutagenesis can be avoided or minimized. Second, the regulation of the corrected gene will be controlled by the endogenous promoter, resulting in a more physiologic regulation of the gene expression.

Finally, diseases with dominant negative mutations can also be addressed, whereas with traditional gene therapy it was not possible. However, the delivery of genome editing tools is challenging. The nuclease components can be delivered as DNA expression cassettes by non-integrating viral vectors, but they can also be administered as RNA or protein complexes by non-viral vectors. Moreover, if the goal is gene correction by HDR, the homologous DNA template must also be provided [285].

While standard gene therapy has many years of clinical experience, the genome editing is still in its infancy. More preclinical studies and carefully designed clinical trials need to be performed to assess the safety of these approaches. The extent of off-target mutations and the immunogenicity of nucleases need to be further investigated [285]. Although this is a relatively new field, more than 40 clinical trials with genome editing tools have been approved. Most of them are *ex vivo* therapies to treat HIV-1 infection or cancer through the modification of T cells [329,330].

2.2.3.1. DSB repair mechanism

2.2.3.1.1. Homology-directed repair (HDR)

HDR repair mechanism can only be activated in the presence of a homologous DNA template (sister chromatid or exogenous DNA). With this system it is possible to introduce the correct copy of a mutated gene or repair a mutation *in situ*. There are several known methods for HDR: canonical HDR, synthesis-dependent strand annealing (SDSA), single-strand annealing (SSA), and break-induced repair (BIR) (Fig. 9). Canonical HDR and SDSA are error-free systems as long as the donor DNA template sequence is correct. With canonical HDR there can be a crossover between the genomic DNA and the donor DNA template. SSA does not require the presence of an external donor DNA, but produces small deletions. On the other hand, BIR is only activated in the presence of one-ended DSBs and can cause chromosomal translocations. It has biological importance, but it is not useful for genome editing [320,322,331,332].

2.2.3.1.2. Non-homologous end joining (NHEJ)

NHEJ repair mechanism does not require a donor DNA template and extensive DNA synthesis, so it is faster than HDR. There are two known methods for NHEJ: canonical NHEJ, and microhomology-mediated end-joining (MMEJ). Both pathways are error-prone, so they can be used to introduce frameshift mutations in coding sequences of aberrantly expressed genes. Canonical NHEJ often causes small insertions or deletions (indels) [320,321,322,332], while MMEJ causes larger indels [333] (Fig. 9).

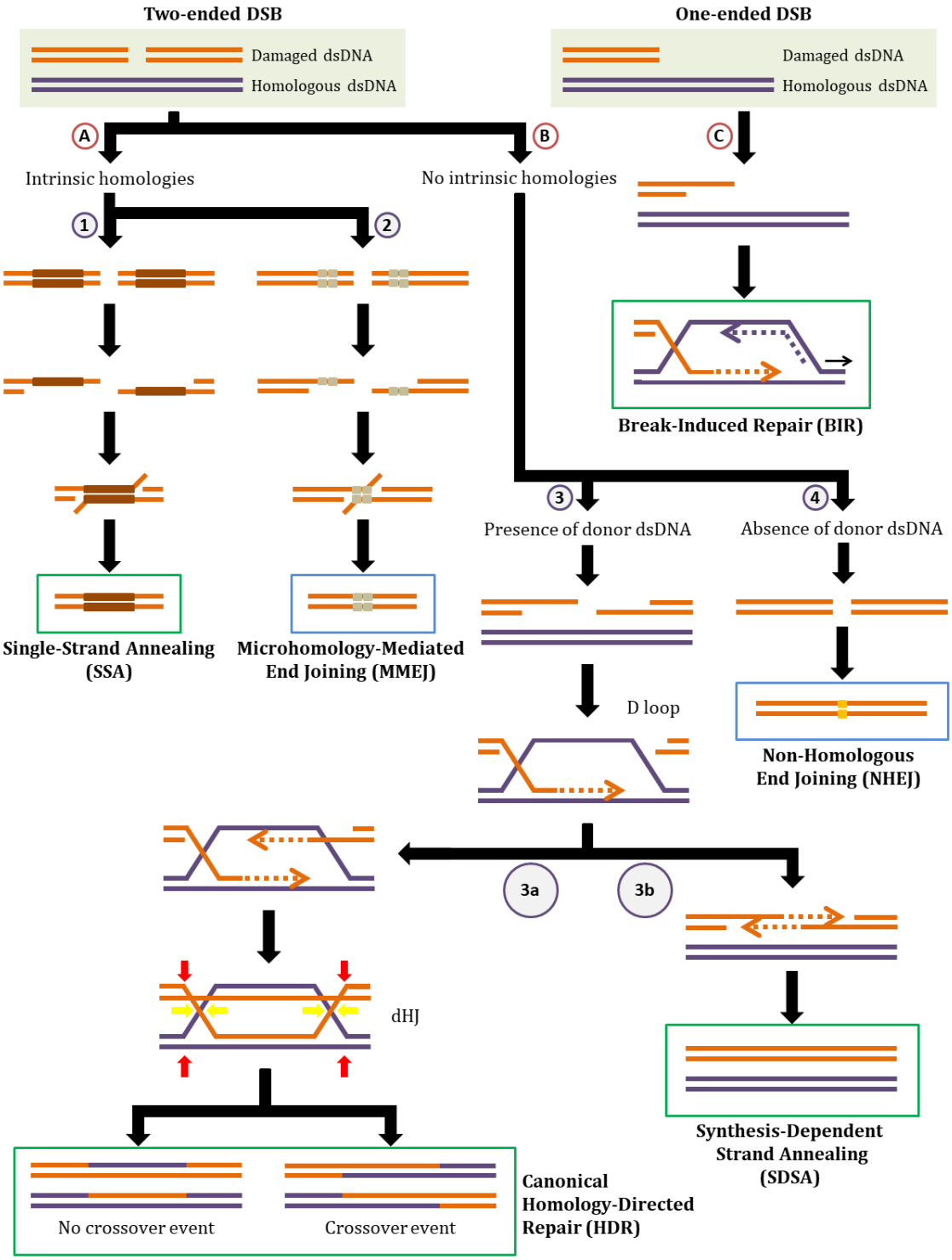

Figure 9. **DSB repair pathways.** The DSB repair pathways can be classified into two global mechanisms: HDR (green squares) and NHEJ (blue squares). **A)** Two-ended DSBs can be resolved by different methods. If the damaged DNA presents intrinsic homologies near the break spot, it can be repaired by two different pathways: **1)** SSA: since the processed 3'-overhangs are complementary, they directly hybridize between each other and the gaps are resolved; **2)** MMEJ: the processed 3'-overhangs anneal through microhomologies, the heterologous flaps are removed and the gaps are filled. **B)** If the damaged DNA does not have intrinsic homologies it can be repaired by different mechanisms depending on the presence/absence of a homologous donor dsDNA. **3)**

Presence of donor DNA: one processed 3'-overhang invades the donor dsDNA and forms a hybrid DNA structure called "displacement loop" (D loop). Using this donor DNA as a template, the invading 3'-end is elongated. This can be resolved in two different ways: **3a)** Canonical HDR: the second 3'-overhang is captured and starts elongating as well, forming a four-stranded branched structure called "double Holliday junction" (dHJ). In the end, the dHJ is resolved by the cleavage of the two junction strands. This resolution can occur in the crossing strands (yellow arrows) or in the non-crossing strands (red arrows). Each of the two junctions can be resolved in either way. If they are resolved dissimilarly, a crossover event will occur. But if both are resolved in the same manner, there will be no crossover. **3b)** SDSA: the invading 3'-end is displaced from the dsDNA template and hybridizes with the non-invading 3'-overhang. Then the non-invading 3'-end is elongated using the invading strand as a template, until the gap is filled. **4)** *Absence of donor DNA*: the DSB can only be resolved by canonical NHEJ: the two DNA extremes are processed to obtain blunt-ended fragments, and then they are directly ligated. **C)** One-ended DSBs can only be resolved by BIR: the processed 3'-overhang invades a complementary dsDNA and acts as a primer in a DNA replication fork [320,321,322,331,332,333].

2.2.3.2. Genome editing tools

2.2.3.2.1. Zinc Finger Nucleases (ZFNs)

ZFNs are the first-generation gene editing tools. They are artificial enzymes generated by the fusion of two domains: the site-specific DNA-binding domain; and the DNA-cleavage domain [334,335,336]. The DNA-binding domain typically contains 3-6 individual Zinc Finger (ZF) units. Each ZF unit selectively recognizes 3 base pairs (bp) of DNA [336,337]. The ZF motifs were discovered in 1985 as part of a transcription factor IIIa in *Xenopus* oocytes [338], and are considered to be the most common type of DNA-binding motifs in eukaryotic transcription factors [339]. The DNA-cleavage domain that is typically used is the *Flavobacterium okeanokoite* type II restriction endonuclease (FokI). Since this enzyme needs to be dimerized to cleave DNA, two ZFN molecules are required to bind to the target site in an appropriate orientation. This means that two different ZFNs have to be generated, one for each side of the DNA target site, with proper spacing (5-7 bp) between them [336,340].

However, this system is expensive and requires specialized expertise to generate artificial proteins consisting of customizable sequence-specific DNA binding domains, each connected to a non-specific nuclease for target cleavage. Moreover, arrays of ZF motifs influence neighbouring ZF specificity, so their design and selection is challenging [336,335] (Table 6).

The first genome editing clinical trial started in 2009. ZFNs were used to disrupt the *CCR5* gene in T cells *ex vivo*. The inactivation of this gene conferred resistance to HIV-1 infection [341]. From then on, several alternative strategies to improve efficacy were developed and tested in clinical trials [329,330,342]. There are also clinical assays for diseases like haemophilia B, β-thalassemia, sickle cell disease (SCD), and mucopolysaccharidosis (MPS) I and II [329,342].

One promising strategy is the targeted introduction of genes downstream of the highly active albumin (ALB) promoter in hepatocytes. Exon 1 of the *ALB* gene encodes a secreted peptide that is cleaved from the final ALB product, so the addition of the cDNA with a

splice acceptor (SA) site into intron 1 allows the creation of a new protein combining the secretory peptide and the protein of interest [312,343,344]. This approach has been clinically tested in haemophilia B [343], MPS-I [345] and MPS-II [346].

2.2.3.2.2. TALE Nucleases (TALENs)

In 2009, it was discovered that the DNA binding domain of bacterial protein called transcription activator-like effector (TALE) could be easily modified [347,348]. This opened the door to the creation of TALENs, the second-generation gene editing tools [349,350]. TALENs are artificial enzymes generated by the fusion of 2 domains: the site-specific DNA-binding domain; and the DNA-cleavage domain. The DNA binding domain consists of a highly conserved sequence from TALE, a bacterial protein that alters transcription in plant cells. This conserved structure is composed of an array of units, each of which has a specific motif that recognizes a single bp in dsDNA [347,348]. TALE was originally discovered in the phytopathogenic *Xanthomonas* bacteria [348]. Like ZFNs, the DNA-cleavage domain is the FokI type II restriction endonuclease. So TALEN modules also need to be designed in pairs to bind opposing DNA target loci, with proper spacing (12-30 bp) between them [348,350].

TALENs can cleave any DNA sequence of interest with relatively high frequency and exhibit better specificity and efficiency than ZFNs. Also, each TALE unit has no influence on the binding specificity of its neighbour. However, the main challenges are the design of a complex molecular cloning for each new DNA target and its low efficiency of genome screening in successfully targeted cells. They are bigger molecules that can target longer gene sequences, but this complicates their delivery. Moreover, like ZFN, the engineering of TALENs is expensive [351] (Table 6).

The first clinical trial with TALENs started in 2016. It consisted of an *ex vivo* therapy based on engineered CAR-T cells [352]. From then on, several alternative strategies have been developed and tested [330,342].

2.2.3.2.3. CRISPR/Cas9

The clustered regularly interspaced short palindromic repeats (CRISPR) were originally discovered in *Escherichia coli* (*E. coli*) in 1987. Scientifics observed the presence of several nucleotide repeats that were flanked by unrelated and non-repetitive short sequences (spacers) [353]. In the following decade, similar sequences were found in other archaea and bacteria [354], although their function was still unknown. In 2002, several well-conserved genes with putative helicase and nuclease domains were located adjacent to CRISPR, which were named CRISPR-associated (Cas) proteins [355,356]. Later, researchers discovered that CRISPR spacers were derived from foreign viruses or plasmids [357,358], so it was determined that CRISPR sequences and Cas proteins took part in prokaryotic adaptive immune response against bacteriophage infection and foreign DNA [359,360,361,362]. Finally, it was revealed that a subset of this sequence-specific immune response was mediated by a

two RNA-structure derived from CRISPR, and that this structure could be combined into a single RNA, which was named guide RNA (gRNA) [363]. Since this mechanism uses molecules encoded in the CRISPR locus and Cas proteins, it was named CRISPR/Cas system.

The revolution in the genome editing field came in 2012 with the demonstration that CRISPR/Cas9 could be efficiently programmed to cleave basically any desired DNA sequence by simply designing a specific gRNA complementary to the target site of interest [363]. With the demonstration that this system could edit mammalian cells, the field of gene editing evolved rapidly, as now researchers were able to design a reliable and effective tool by only changing the gRNA sequence [364,365,366,367,368]. CRISPR/Cas9 technology became the third-generation gene editing tool.

CRISPR/Cas9 has several important advantages over other gene editing technologies. First, the Cas9 enzyme does not determine the target specificity, so there is no need of protein engineering. Second, the target specificity is determined by the base pairing of a small gRNA, so this platform can be easily mass produced for multiple applications. Third, this system allows the simultaneous targeting of multiple loci. In conclusion, this platform is simple, reliable, efficient, scalable and affordable [365,369] (Table 6).

Gene editing tool	ZFN	TALEN	CRISPR/Cas
Targeting efficacy	Poor	Moderate	High
Cost	Very expensive	Expensive	Cheap
Risk of off-target mutagenesis	Variable	Low	Moderate
Parameter to design	Protein	Protein	RNA

Table 6. Comparison of the properties of the current genome editing tools [370]

2.2.3.2.3.1. Classification

There are multiple CRISPR/Cas systems, which are divided into two classes based on Cas protein composition differences and sequence divergence between the effector modules [371]. Class 1 systems have multiple Cas proteins as effectors, and are divided in 3 types (I, III, IV). Class 2 systems have a single effector Cas protein, and are divided into 3 types (II, V, VI) [371]. Each type is divided in multiple subtypes. All systems have in common that they use the CRISPR locus as an adaptive immune system against foreign mobile genetic elements (MGEs), but they present many differences in their mechanism of action and the type of molecular effectors [372]. Moreover, the different systems are not uniformly distributed among bacterial and archaeal phylogeny. While the majority of archaea have CRISPR/Cas loci (85,2%), they are less frequent in bacteria (42,3%). However, class 2 systems are nearly exclusive to bacteria [371].

Class 2 type II CRISPR/Cas systems use a single Cas9 protein as effector. There are several Cas9 orthologues with different properties, which are classified in three subtypes (A, B, C).

The most frequently used is the type II-A CRISPR/Cas9 system (class 2, type II, subtype A), which uses a single Cas9 protein to produce DSBs in dsDNA, and ribonuclease (RNase) III for the maturation of gRNA [371,372]. This was the first engineered CRISPR/Cas system, so there is extensive knowledge about it [363]. For this reason, all subsequent information about the CRISPR/Cas mechanism will refer to this particular system, otherwise specified.

2.2.3.2.3.2. CRISPR/Cas9 elements

gRNA

The gRNA is the CRISPR/Cas9 element tasked with DNA sequence recognition. This RNA structure interacts with the Cas9 enzyme and directs its DNA cleavage to a specific sequence [363,364]. gRNA is formed by 2 small molecules, a target-recognizing RNA (crRNA), and an auxiliary non-coding trans-activating RNA (tracrRNA). Although the natural gRNA is formed by two hybridized RNA molecules, a single gRNA can be constructed by fusing crRNA and tracrRNA [363] (Fig. 10).

The crRNA contains a unique 20 bp sequence designed to complement the foreign DNA. These 20 bp can be any sequence, but they must be adjacent to a short DNA sequence (2-5 nucleotides) upstream called "Protospacer Adjacent Motif" (PAM). This PAM is essential for the compatibility with the Cas9 protein used [363,364,365,367,373,374,375]. For this reason, before designing a gRNA the genome needs to be screened for the presence of PAMs, as only the sequences downstream of PAM can be targeted by CRISPR/Cas9. It is important to note that the gRNA has to be complementary to the 20bp downstream PAM, but not to PAM itself.

Cas9 enzyme

The Cas9 protein is the CRISPR/Cas9 element tasked with the cleavage of the DNA sequence. The gRNA positions the Cas9 enzyme in a specific locus, and the protein catalyses the DSB. But Cas9 also interacts with the DNA and needs the presence of a PAM sequence to perform the DNA cleavage. In absence of PAM recognition, the Cas9 enzyme does not function properly. PAM is different depending on the Cas9 bacterial origin. The canonical Cas9 used is from *Streptococcus pyogenes* (*S. pyogenes*) (SpCas9), which has NGG as PAM [363,364,365,375].

SpCas9 has a length of 1368 amino acids and is divided in two lobes: the nuclease (NUC) lobe and the α-helical recognition (REC) lobe. The NUC lobe is divided in two domains: the HNH domain, which cleaves the complementary DNA strand; and the RuvC domain, which interacts with PAM and cleaves the non-complementary DNA strand. The REC lobe is responsible of sensing nucleic acids, regulating the conformational transitions and site locking the HNH domain [363,376,377]. SpCas9 produces a 1-bp staggered end with 5'-overhangs due to the post-cleavage trimming activity of the RuvC domain [378] (Fig. 10).

2.2.3.2.3.3. CRISPR locus

There are several types of CRISPR loci depending on the CRISPR/Cas system that they encode and their prokaryotic origin. They have common elements, like the CRISPR array and the Cas gene cluster, but their sequence, regulatory elements and gene organization varies greatly. The same prokaryotic species can present several different CRISPR loci [354,355,374] (Fig. 10).

CRISPR array

The CRISPR array consists of several repeat sequences (23-47 bp) interspaced by spacers (21-72 bp). The repeat sequences are highly conserved within a CRISPR locus, but they vary across different CRISPR/Cas systems and prokaryote species. They are often palindromic sequences [354,355,356,374]. The spacer sequences are highly variable but have similar length within the same array. New spacers are introduced during a viral infection to acquire memory and block subsequent invasions [357,360]. A leader sequence located upstream permits the transcription of the CRISPR array, which results in a long pre-crRNA [355]. Later this pre-crRNA is divided in units, each of which is a mature crRNA formed by a spacer flanked by two partial repeats. The repeats are common sequences present in all crRNAs, while the spacers are unique sequences complementary to specific fragments of foreign MGEs [361,374]. In the case of type II CRISPR/Cas9 system, a fragment of the repeat sequence is complementary to tracrRNA [370,379] (Fig. 10).

As new spacers are introduced, old spacers are deleted. Spacers located at the end of the array are preferentially deleted because they are more likely to be ancestral spacers from historical infections. Perhaps this allows the size limitation of the CRISPR array so that the relative size does not increase to a detrimental level. In conclusion, the CRISPR array undergoes dynamic and rapid turnover on evolutionary time scales [374].

***tracrRNA* gene**

The tracrRNA molecules are encoded in an independent gene inside the CRISPR/Cas locus. This tracrRNA hybridizes with the repeats present in the pre-crRNA and plays an important role in its maturation. This element is not necessary in all CRISPR/Cas systems, but is essential for type II CRISPR/Cas9 [363,370,379,380] (Fig. 10).

Cluster of *cas* genes

Inside the CRISPR/Cas locus there is a cluster of different *cas* genes, which encode a large and heterogeneous family of proteins that carry functional domains typical of nucleases, helicases, polymerases and nucleotide-binding proteins [355,356,359,374,370,379] (Fig. 10). All *cas* genes can be subdivided into four distinct functional modules: the adaptation module, which is responsible of the acquisition of new spacers; the expression processing module, which is responsible of pre-crRNA processing; the effector module, which is involved in target recognition and nucleic acid cleavage; and the ancillary module, which is a diffuse collection of CRISPR-linked genes with a predicted role in the CRISPR/Cas systems. The same Cas protein can belong to more than one module, as it can have more than one function [371].

In the case of type II-A CRISPR/Cas9 systems the adaptation module is composed mainly of Cas1, Cas2 and Csn2. RNase III is the main protein of the expression processing module. Cas9 is the only protein of the effector module, but this enzyme also plays a role in pre-crRNA maturation (expression processing module), and in PAM selection during adaptation (adaptation module). Finally, Csn2 is predicted to have a helper role yet unidentified, so it can be added in the ancillary module [371].

2.2.3.2.3.4. Adaptive immune mechanism of CRISPR/Cas9

CRISPR/Cas9 adaptive immune response is mediated by an RNA-directed DNA endonuclease (Cas9) that introduces a DSB to foreign MGEs [363]. This mediated adaptive immunity is divided into three phases [379,370] (Fig. 10):

- **Adaptation**: During this phase the prokaryotic cells acquire genetic memory of viral infections. When a phage infects host bacteria, the invading DNA is processed and small DNA fragments are obtained (protospacers). The selection of specific protospacers and their orientation depends on the detection of PAM sequences in the viral genome by Cas9. Those protospacers are then incorporated into the CRISPR array as new spacers by Cas1, Cas2 and Csn2. In the end, the CRISPR array is composed of different spacers that are complementary to the viral DNA of a different phage infection [360,373,381,382,383,384,385,386] (Fig. 10).

- **Expression**: Both *tracrRNA* gene and CRISPR array are transcribed, and Cas9 enzyme is synthesized. Multiple tracrRNAs combine with the pre-crRNA and form dual RNA hybrids, which are stabilized by Cas9. These RNA duplex conformations are recognized by RNase III, which cuts the pre-crRNA in monomers. Each separated unit is a gRNA conformed by a tracrRNA and a mature crRNA that recognizes a different viral sequence. These gRNAs are coupled to Cas9 enzymes, so the effector molecules are ribonucleoproteins (RNPs). The interaction of these elements induces the activation of Cas9 through a conformational change [361,380,387] (Fig. 10).

- **Interference**: When a phage invades the prokaryote a second time, it already has a mechanism to detect it and stop the infection. Cas9 initiates the target DNA search by probing for a proper PAM sequence before interrogating the flanking DNA for potential gRNA complementarity. When a proper target is located Cas9 triggers local DNA melting, and the gRNA performs a strand invasion and generates an R-loop (RNA-DNA hybrid + displaced DNA strand). The interaction with the correct DNA target triggers specific conformational changes in Cas9 that allows the stabilization and activation of the enzyme. Finally, Cas9 produces DSB at 3-8 nucleotides upstream PAM [362,363,364,375,376, 377] (Fig. 10).

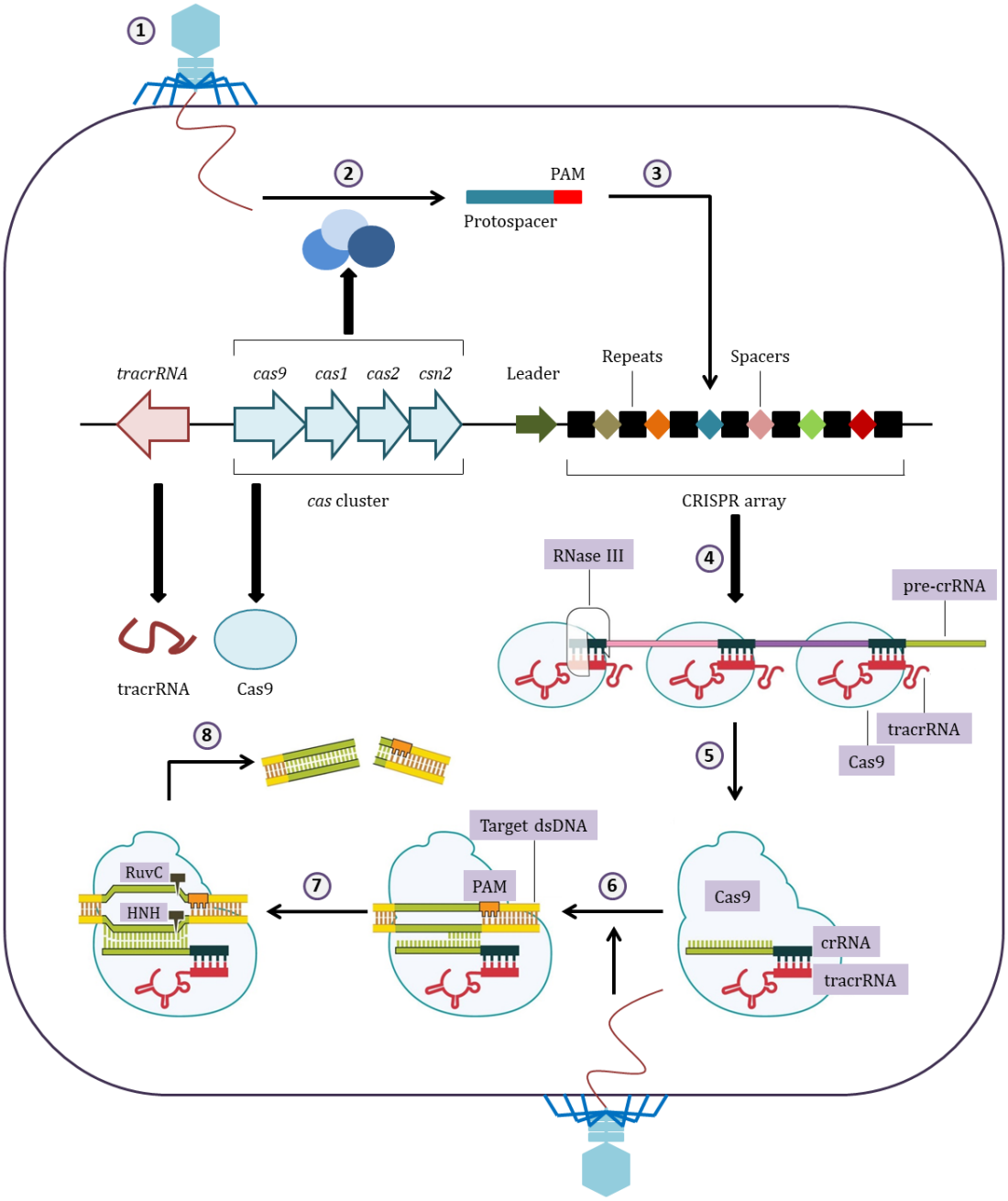

Figure 10. **CRISPR/Cas9 structure and action mechanism.** The adaptive immune mechanism of CRISPR/Cas9 can be divided in 3 steps. Adaptation: When a viral infection happens (1), the Cas enzymes process the viral genome and generate fragments called protospacers (2). These structures are incorporated to the CRISPR array as new spacers (3). Expression: The transcription of the CRISPR array produces a pre-crRNA (4). Cas9 and tracrRNA bind to this pre-crRNA and promote its maturation through the action of RNase III (5). Interference: When a virus infects the host cell a second time, there are already CRISPR/Cas9 molecules that recognize its genome (6). Cas9 interacts with PAM, and the gRNA (crRNA + tracrRNA) binds to the target sequence (7). Then Cas9 cuts the viral dsDNA: HNH domain cleaves the complementary DNA strand; and RuvC domain cleaves the non-complementary DNA strand (8) [368].

2.2.3.2.3.5. Challenges of CRISPR/Cas9 technology

Avoid off-target mutations

Off-target genome editing refers to DNA modifications at unintended and non-specific sites. While CRISPR/Cas9 is highly specific in bacterial cells, in mammalian cells this system has tolerance for sequence mismatch, so both identical and highly homologous DNA sequences can be cleaved [388]. Compared to other gene editing tools, it has a higher potential for off-target effects in human cells. The production of DSB in undesired loci could potentially lead to genotoxicity, genome instability, gene function disruption, epigenetic alterations or carcinogenesis. For this reason, the improvement of CRISPR/Cas9 specificity has become top priority [388,389,390].

It is important to note that for clinical applications a risk-benefit ratio has to be assessed. The identification of loci susceptible to off-target effects in each therapy is important. Obviously, off-target cleavage in proto-oncogenes or tumour suppressor genes must be avoided, but the disruption of certain genes could be tolerated depending on the target tissue. For example, if the goal is to edit a gene in the hematopoietic lineage, the possible off-target modification of a gene that is only expressed in muscle could have no deleterious effects. Moreover, off-target modifications could be tolerated if they are random (no sequence bias) and the levels are equal or below the random mutations frequencies. In conclusion, the presence of off-target effects does not necessarily preclude clinical application, but it is important to study each genome editing strategy separately. The treatment risk must be weighed with the disease prognosis and the potential benefit [330,391].

The assessment of CRISPR/Cas specificity can be conducted by either biased or unbiased methods [392,393,394]. For a biased off-target analysis, several bioinformatics tools have been developed to predict potential non-specific targets with PAM compatibility [372,388,390,394]. Then these loci are screened by methods like mismatch detection nuclease assay [395,396] or NGS [396,397,398,399,400]. In contrast, unbiased off-target analyses have been developed to search the whole genome for mutations in non-specific targets [372,394]. This can be done by many different methods, such as WGS [399,344], WES [401], Guide-Seq [402], Digenome-Seq [403], BLESS [404], or LAM-HTGTS [405].

There are several factors that might affect the specificity of gene editing. First, the intrinsic Cas9 specificity may determine the relative importance of each nucleotide position of the target sequence, meaning that Cas9 determines which nucleotide position of the 20 bp is more prone to accept a wrong match. Second, the structure and composition of gRNA can affect the level of off-target cleavage. Third, the specificity also depends on the abundance of effective CRISPR/Cas9 complexes relative to the genomic target concentration [329,406].

To improve the intrinsic Cas9 specificity several variants have been generated. They are engineered to reduce non-specific interactions with DNA [399,407,408,409]. Another strategy is the use of Cas9 nickases (nCas9), which produce single-strand breaks (SSBs) instead of DSBs. The employment of two nCas9 that target nearby sequences allows the production of a DSB with sticky ends. This system can significantly avoid off-target cleavage without sacrificing genome editing efficiency [401,410]. Another alternative is Cas12a (Cpf1), a simpler Cas nuclease used by *Prevotella* and *Francisella*. Cas12a is an effector of a class 2 type V

CRISPR/Cas system. It generates sticky ends and presents high specificity and HDR efficiency. Moreover, it does not need tracrRNA so the design process is simpler [400,411,412].

Some methods to refine gRNAs are the selection of unique target sequences that lack homology to other genomic regions and the engineering of the gRNA structure [401,413,414]. It has been described that mismatches at the proximal 5' region of gRNA are better tolerated than those at the 3' region [369,388,389]. Additionally, it has been reported that the use of truncated and less-active gRNAs decreased off-target effects [396,397].

In addition, the concentration of CRISPR/Cas9 delivered to cells should be adjusted to the minimum necessary for efficient cleavage, as the presence of surplus molecules increases the chance of off-target cleavage [329,388,397]. One way to have more control of the CRISPR/Cas9 molecules delivered is the administration of the RNP instead of the plasmid that encodes the elements. It has been demonstrated that direct delivery of recombinant Cas9 protein and *in vitro* transcribed gRNA presents reduced off-target effects [415]. Another way to control the CRISPR/Cas9 concentration is the use of inhibitors [416].

Increase HDR efficiency

The repair of DSBs varies greatly between cell types and cell status, but in most cases NHEJ is more active than HDR. NHEJ is active in all stages of the cell cycle, while HDR functions primarily in the S/G2 stage and is therefore largely restricted to actively dividing cells [322,319,328]. However, a recent report showed that post-mitotic neurons could be edited through HDR using AAV vectors to deliver the donor DNA template [417]. Even then, NHEJ is more efficient and has the advantage of not needing the supply of a donor DNA. But gene editing through NHEJ has limited applications, as it is mainly useful to disrupt gene expression. On the other hand, the ability to introduce a donor DNA at a specific locus can be useful in many applications, so the increase of HDR efficiency is needed.

There are multiple strategies to favour HDR, such as inhibiting NHEJ [418,419,420,421,422,423,424], enhancing HDR [418,419,422], or timing the CRISPR/Cas9 delivery by synchronizing cell cycle [398]. A recent report claimed that performing a “cold shock” in cultured cells increased HDR efficiency [425]. Another option is to discard the HDR mechanism altogether and introduce a donor DNA in a HDR-independent manner. This can be done using homology-independent targeted insertion (HITI), which consists of the introduction of a donor DNA template without homology arms through NHEJ. Since NHEJ is highly efficient, the introduction of a donor DNA by this method is more efficient than with HDR [426,427,428].

Another challenge is the design of the donor DNA template. Although there are common mutations in certain diseases, there is mutational variability amongst patients. For this reason, a strategy that permits the correction of all existing mutations needs to be designed to allow an effective scaling production in the future [429]. Consequently, more often than not it is necessary to introduce the whole coding sequence (cDNA) of a gene to encompass all possible mutations. Linear DNA is more prone to be introduced by HDR than circular DNA [430]. In the case of plasmid delivery, the addition of CRISPR target sites flanking the donor template can enhance the linearization of the donor DNA, and thus increase HDR efficiency [431]. The size of the homology arms (HAs) is a controversial point. Some reports use long HAs [344,369,417,427,432,433,434,435,436,437,438], while other works claim that

only a small homology is needed [418,425,430,439,440,441,442,443,444]. There is also evidence that ssDNA donor presents higher HDR efficiency than dsDNA [418,425,439,441,443,444]. Since Cas9 first releases the 3' end of the non-target strand, the use of ssDNA donors with one small HA complementary to this strand could favour HDR [441].

Overcome target restriction due to PAM

Since PAM is an indispensable feature for CRISPR/Cas activity, the genomic targets that can be edited are limited to loci that contain these PAM sequences. On a positive note, there are multiple available Cas orthologues that detect different PAMs. But even then, it is possible that the desired target does not have it [372].

A possible way to overcome this is by engineering Cas9 enzymes to make them capable of recognizing a wide-range of PAM sequences [445,446]. Cas12a variants with alternative PAM recognition have also been developed [447].

Avoid immunogenicity

The CRISPR/Cas9 system has a prokaryotic origin, so the *in vivo* delivery can trigger innate and adaptive immune responses.

Plasmid DNA, mRNA Cas9 or gRNAs can be recognized by the innate immune system. These nucleic acids present PAMPs that can be recognized by PRRs. This interaction activates several signal transduction pathways that promote pro-inflammatory cytokine production [448,449]. The chemical modification of the nucleic acids and the optimization of its sequence could help temper this response. Particularly, the immune system recognizes the 5' terminal phosphate group, so the modification of the phosphate into a 5' hydroxyl group has been suggested [450,449].

The Cas enzymes trigger adaptive immune responses, which means that modified cells could be eliminated by CD8+ T cytotoxic cells. This can be avoided by immune suppression, but it is a short-term solution. An option to reduce immunogenicity is the engineering of Cas9 proteins to alter reactive epitopes without compromising their functionality [451]. Moreover, the transient presence of Cas9 could prevent this adverse effect, so the RNA or RNP delivery approaches are better.

Additionally, during the first exposure the organism generates a memory T cell population that can trigger a rapid humoral and cellular response upon a second exposition. Since *S. pyogenes* is prevalent in human environments, it is likely that many potential patients present pre-existing immunity to SpCas9. Indeed, a study evidenced that 58% of adult humans carried antibodies against SpCas9, and 67% had anti-SpCas9 T cells [452]. Other studies showed that the systemic delivery of SpCas9 by viral vectors resulted in the development of anti-SpCas9 antibodies and cellular immunity in animals [453,451]. It may be interesting to seek alternative Cas9 enzymes from organisms less contagious to humans [372].

Optimize the delivery system

The CRISPR/Cas9 system can be delivered in three different formats, each of which has its advantages and limitations (Table 7):

- DNA-based approach

Plasmids or linear DNA fragments that encode the Cas9 protein and the gRNA can be used. This strategy produces a stable expression and is more cost-effective than the others. However, it has several limitations. First, CRISPR/Cas9 expression is relatively long lasting, which increases the chance of off-target cleavage and immunogenicity. Second, there is a risk of random integration of the plasmid into the host genome. Third, the plasmid needs to reach the nucleus, which is challenging [454].

Viral vectors can be used to introduce the cassette expression of Cas9 and gRNA. ADV vectors [453,455,456] and LV vectors [457] have been successfully used in mice, but they may elicit immune responses in the host. AAV vectors have the best safety profile, but they have a limited packaging capacity, so the delivery of a large gene like Cas9 is challenging [299]. One way to overcome this is the packaging of Cas9 and gRNA in separate AAV vectors [396,424,435,437,458,459,460]. Alternatively, SpCas9 can be split into two AAV vectors [434,451]. Another option is the use of Cas9 orthologues with a smaller size, like Cas9 from *Campylobacter jejuni* (CjCas9) [461], *Staphylococcus aureus* (SaCas9) [460,462,463,464] , or *Neisseria meningitidis* (NmCas9) [465].

The direct delivery of a CRISPR/Cas plasmid through hydrodynamic injection [439,466] or electroporation [467,468] has been successfully performed in animal models. There is also evidence of successful non-viral vector delivery of CRISPR/Cas DNA [469,470,471,472,473].

- RNA-based approach

The nuclease complex can be delivered as Cas9 mRNA and gRNA. SpCas9 mRNA has a length of 4300 nucleotides approximately, while gRNA has a length of 100 nucleotides. Both molecules have negative charge. Since they need to be delivered together into target cells, the carriers need to be able to encapsulate both long and short RNA at the same time [454,474]. This strategy avoids the problem of nuclear entry of the Cas9 gene, because the RNP complexes are formed directly in the cytosol, and this also permits an earlier onset of gene editing. Moreover, RNA has a short half-life (up to 24 hours) so the risk of off-target effects and immunogenicity is lower. However, this transient expression may lead to low efficiency in gene editing. But the biggest challenge is the poor stability of RNA compared to DNA [454].

Various methods have been developed to improve RNA stability. For Cas9 mRNA, it has been tested the use of Cap analogues to improve mRNA co-transcriptional capping, or the introduction of chemical modifications to enhance resistance to RNases and reduce immunogenicity [475,476]. For gRNA, chemical modifications have reported to significantly enhance the genome editing efficiency and reduce immunogenicity [477,478].

CRISPR/Cas RNAs has been successfully delivered with several NPs [413,436,478,479,480].

- RNP-based approach

The direct delivery of the RNP (Cas9 protein and gRNA) has also been tested. SpCas9 protein has a length of 1368 amino acids (160 kDa) and is positively charged. Both SpCas9 and gRNA need to be delivered together into target cells, so the carriers need to be able to encapsulate both protein and short RNA at the same time [454, 474]. This strategy allows a fast gene editing, as there is no need of transcription or translation. Moreover, the complex has a great stability. The CRISPR/Cas9 effect is also transient, so the risks of off-target effects and immunogenicity are lower. The main challenges of this approach are the delivery of a large protein (160 kDa), the positive charge of Cas9, and the super-negative charge of gRNAs [299]. Also, Cas9 needs to be carefully purified before its delivery to avoid contamination with bacterial endotoxin, which is a costly procedure. Another important point is that unlike the other delivery platforms, which mainly induce innate immune responses, the Cas9 protein can induce adaptive immune responses [452,451,454].

Several types of NPs have been successful in the delivery of CRISPR/Cas RNP [481,482,483,484,485,486]. The microinjection of zygotes [444] or the direct delivery of the RNP [487] has also been tested.

Delivery mode	Advantages	Disadvantages
Plasmid DNA	Low cost High stability	Low efficiency Delayed onset Risk of insertional mutagenesis
RNA	No insertional mutagenesis Transient duration Low off-target effects Rapid onset	Poor stability
RNP	High stability No insertional mutagenesis Transient duration Low off-target effects Rapid onset	High cost Risk of bacterial endotoxin contamination

Table 7. Comparison of the CRISPR/Cas delivery strategies [299,454]

2.2.3.2.3.6. Applications

Multiple therapeutic strategies based on CRISPR/Cas9 gene editing have been reported in animal models for treating viral infections [466,480,488], hereditary diseases [426,435,436,437,439,442,456,459,460,424,462,463,464,465,467,468,469,478,480,484,485,486,489], and cancers [438,472,483,487,490]. At present there are several ongoing clinical trials for diseases like cancer, HIV-1 infection, neurofibromatosis type 1, β-thalassemia, LCA10 and SCD [329,330].

This system has also been used to screen the functionality of genes and to create cellular and animal models of diseases. It can also be used as an antimicrobial and antiviral [370,429,491].

Apart from gene editing, the CRISPR/Cas9 system has also been engineered to present alternative uses. The Cas9 protein can be modified to be catalytically inactive (dead Cas9, dCas9) and can be fused with other molecules. One way to use dCas9 is to target gene promoters and prevent the interaction of the transcriptional machinery. With this method the expression of a certain gene can be inhibited [492]. Another reported approach is the use of dCas9 fused with transcription activators or repressors. This system allows the regulation of gene expression by recruiting epigenetic regulators [493]. Another reported strategy is the correction of mutations at a single-base level without inducing DSBs. The dCas9 proteins are fused with a deaminase to obtain "base editors", which catalyse transitions of nucleic acid bases and are able to correct point mutations [445,494,495,496]. Alternatively, dCas9 can be fused to a reverse transcriptase and programmed with a prime editing gRNA (pegRNA) that both specifies the target site and encodes the desired edit [497]. Finally, dCas9 can be fused with fluorescent proteins to visualize nucleic acids [498,499].

Recently, other types of CRISPR/Cas systems have also been tested. For example, type VI systems (class 2), which have Cas13 enzymes as effectors, are able to target RNA. These nucleases can be used for RNA interference or imaging [500].

3. HYPOTHESIS AND OBJECTIVES

3.1. Hypothesis

The hypothesis of this study is that long-term TP expression can be achieved in liver cells of the murine MNGIE model by using the CRISPR/Cas9 gene editing tool and rAAV2/8 vectors carrying DNA templates containing the *TYMP* cDNA. These templates will be designed to be integrated in a region downstream a genomic promoter, so that the *TYMP* cDNA expression will be under its control. The CRISPR/Cas9 molecules will increase the chances of integration by producing DSBs at the desired location. Two different loci will be targeted, the endogenous *Tymp* locus, and the *Alb* locus, as its promoter is highly active in hepatocytes.

Two distinct strategies will be tested. In the first one, CRISPR/Cas9 RNAs will be delivered by NPs. This will allow a high but transient expression. In the second approach, CRISPR/Cas9 will be delivered as DNA inside rAAV2/8 vectors.

Successful gene editing will restore the nucleoside homeostasis in the MNGIE murine model. If effective, the same or similar approaches will be useful for other monogenic diseases.

3.2. Objectives

The main objectives of this work are the following:

1. Design and select a specific gRNA for each locus. Different gRNAs will be screened and those with highest gene editing efficiency will be chosen.

2. Design and generate suitable DNA templates for each locus. These templates will contain the *TYMP* cDNA and other elements that enable its expression if the gene editing is successful.

3. Test the gene editing strategies *in vitro* to check if the DNA templates are correctly inserted at the desired location. Moreover, we will assess that the *TYMP* cDNA is transcribed and translated into a functional enzyme.

4. Select the NP formulation that has the highest delivery efficiency in hepatocytes. We will test NPs of two different sources: PNPs of GEMAT-IQS, and LNPs of Acuitas Therapeutics®. Both their *in vitro* and *in vivo* efficiencies will be compared.

5. Determine if this gene therapy approach is efficient and therapeutically useful in the MNGIE mice model. rAAV2/8 vectors and NPs will be injected to animals and the gene editing efficiency will be assessed by different methods:

    - Thd and dUrd plasma levels will be monitored. A significant decrease of their levels will indicate biochemical correction.

- Livers will be collected and at the end of experimentation and the successful insertion of TYMP cDNA will be assessed. The presence of TP mRNA and functional TP protein will also be determined.
- In mice of the *Alb* locus gene-editing group, the TP activity will also be assessed in plasma. Correctly edited hepatocytes are expected to produce a hybrid Alb-hTP enzyme that can be secreted.

6. Compare the results obtained in both loci. We will determine which locus is the most suitable candidate to be used in the future.
7. Compare the two gene editing approaches. The use of NPs as an alternative to using only rAAV2/8 vectors will be evaluated.

4. MATERIALS AND METHODS

4.1. Gene editing tools: design and generation

4.1.1. CRISPR/Cas9

In this work the CRISPR/Cas9 platform was delivered as RNA molecules. Several gRNAs were designed and custom-generated, but the same Cas9 mRNA (*L-7606, Trilink Biotechnologies*) was used in all experiments. This mRNA is capped (with a proprietary method), polyadenylated and optimized for mammalian systems. The encoded Cas9 contains two NLS (N- and C-terminal), which increases the frequency of delivery to the nucleus. It mimics a fully processed mature mRNA and exhibits high editing efficiency.

4.1.1.1. Design of gRNAs

gRNAs are 100 nucleotide-long sequences composed by a common structure and a variable region (20 nucleotides) that determines the target specificity (Fig. 11). Several gRNAs were designed for each locus. The first step was the selection of potential CRISPR/Cas9 targets in the desired genomic location. This was done with the bioinformatics tool developed by *Zhang Lab* (*Massachusetts Institute of Technology (MIT), Cambridge, USA*). This tool automatically searched PAM motifs within the target genomic fragment. It also predicted the potential off-target cleavage of each match based on the frequency of similar sequences with PAM compatibility in other genomic loci. In the end, a list of all CRISPR/Cas9 potential targets was obtained, sorted by the putative risk of off-target effects. The first 3-5 targets of each locus were selected. All these gRNAs were generated *in vitro*, and their editing efficacy was evaluated in cell culture. For each locus, the gRNA with better performance was selected and used in all subsequent experiments.

- ***Tymp* locus:** Initially, three gRNAs targeting intron 1 were generated (Tymp1, Tymp2, and Tymp3). Later, the editing strategy changed and two additional gRNAs targeting intron 2 were generated (Tymp4 and Tymp5) (Table 8).
- ***Alb* locus:** Three gRNAs targeting intron 1 were generated (Alb1, Alb2, and Alb3) (Table 8).

4.1.1.2. Generation of gRNAs

The gRNAs used in preliminary experiments were generated with the commercial kit *Precision gRNA Synthesis Kit* (*A29377, Invitrogen*), following the manufacturer's instructions. This methodology required custom-designed specific primers (Table S1), which were purchased from *Thermo Fisher Scientific* (*Waltham, USA*). These primers were used to generate a DNA template that contained a promoter and the coding sequence of the custom gRNA. Then, these DNA fragments were transcribed *in vitro* and the resulting gRNAs were purified. To ascertain the quality of the resulting gRNAs, a 2% agarose gel

electrophoresis was performed (detailed on section 4.7.6). The presence of a single 100 bp band confirmed a pure gRNA yield.

For *in vivo* experiments, synthetic gRNAs were purchased from *Axolabs* (*Kulmbach, Germany*). These gRNAs have chemical modifications to enhance stability and efficacy (Fig. 11).

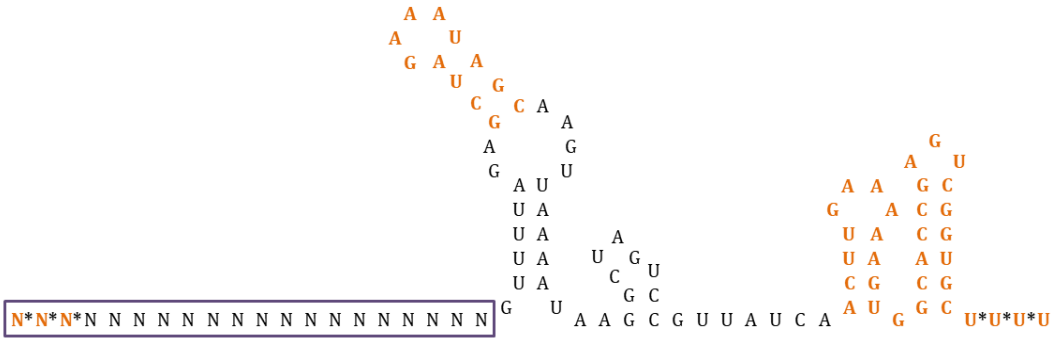

Figure 11. Synthetic gRNA structure with chemical modifications. The framed nucleotides represent the variable region that determines the gRNA specificity. The corresponding gRNA target sequences are detailed in table 8. N: any ribonucleotide (A/U/G/C); asterisk (*): phosphorothioate bond; **N**: ribonucleotide with a 2'-O-methylation modification.

Locus	gRNA	Genomic target sequence (5'-3')	gRNA target sequence (5'-3')
Tymp	Tymp1	(+) GAATAGGTACTAGACGTGGG**CGG**	GAATAGGTACTAGACGTGGG
	Tymp2	(+) AATTGAATAGGTACTAGACG**TGG**	AATTGAATAGGTACTAGACG
	Tymp3	(+) AATAGGTACTAGACGTGGGC**GGG**	AATAGGTACTAGACGTGGGC
	Tymp4	(+) AGCTGCCCCATACGCACCTC**AGG**	AGCTGCCCCATACGCACCTC
	Tymp5	(-) **CCC**ATACGCACCTCAGGGATCCA	TGGATCCCTGAGGTGCGTAT
Alb	Alb1	(+) TTCCTGTAACGATCGGGAAC**TGG**	TTCCTGTAACGATCGGGAAC
	Alb2	(-) **CCT**GCTCGACCATGCTATACTAA	TTAGTATAGCATGGTCGAGC
	Alb3	(-) **CCT**GTAACGATCGGGAACTGGCA	TGCCAGTTCCCGATCGTTAC

Table 8. Characteristics of the gRNAs generated. (+): sense strand; (-): antisense strand; PAM in bold and underlined.

4.1.2. Plasmids

Plasmids basically have two distinct structures. The first one is the plasmid backbone, which contains the elements necessary for plasmid amplification in bacterial culture. These basic elements are the replication origin (Ori) and an antibiotic resistance gene. The second structure is the insert, whose content varies depending on the application.

4.1.2.1. Design of plasmid DNA templates

The insert of these plasmids was the sequence of DNA templates. Over time, different DNA templates were designed and generated. Most of them had some common structures:

- **cDNA**: It corresponds to the coding sequence of a gene. It is a DNA synthesized from the reverse transcription of a mature mRNA. For this reason, cDNAs only contain the sequence of exons.

  In initial experiments, the enhanced green fluorescence protein (*eGFP*) cDNA was used to easily detect edited cells. The human TYMP (h*TYMP*) and murine *Tymp* (m*Tymp*) cDNAs were used to determine the efficiency of the therapy (Table S2).

- **SA**: It is a small motif that marks the limit between introns and exons. The transcription of a gene produces an immature mRNA (precursor mRNA, pre-mRNA) that contains both introns and exons. This pre-mRNA must be processed to become a mature mRNA that can be translated into a protein. RNA splicing is one of the modifications performed in mRNA maturation. During this process introns are removed and exons are coupled together to generate the coding sequence. SAs are the motifs that mark the sites where RNA splicing must be performed. In the present work, DNA templates are introduced inside introns, so the inclusion of SAs assures that the cellular splicing machinery identifies the cDNA inserted as a coding sequence and not a continuation of the intron. This way, the mature mRNA generated from the edited gene contains the cDNA.

  The DNA templates generated contained an endogenous SA from the murine *Alb* gene. The terminal 20 nucleotides of intron 1 were cloned at the 5'-end of cDNAs (Table S2).

- **Polyadenylation signal (pA)**: It is a sequence that triggers the polyadenylation of the transcribed mRNA. Another modification introduced during mRNA maturation is the addition of multiple adenosines (poly-A) at the 3'-end. The cellular proteins detect the pA in a pre-mRNA and start the synthesis of a poly-A tail. This structure is important for the nuclear export, translation, and stability of mRNA.

  The DNA templates generated contained the pA from simian virus 40 (SV40). This sequence was cloned at the 3'-end of cDNAs (Table S2). It includes the motif AAUAAA, which promotes both polyadenylation and termination of transcription. It is used in many mammalian expression plasmids.

- **HAs**: They are sequences that flank the cDNA that needs to be inserted. These sequences are homologous to the edited locus (Table S2).

DNA templates would ultimately be carried by rAAV2/8 vectors in the *in vivo* experiments. But before that, it was necessary to test their efficiency and optimize their structure *in vitro*. For this reason, DNA templates were delivered as plasmids in the initial experiments. This had some advantages, as plasmids are easily engineered and can be rapidly amplified. However, they present additional sequences not needed for gene editing. Also, they are circular DNA molecules so the insertion of the templates through

NHEJ was challenging. Ideally, DNA templates had to be delivered as lineal molecules without unneeded sequences. This way the experimental settings would better mimic the *in vivo* conditions, and it would allow the assessment of NHEJ efficiency.

4.1.2.1.1. First generation DNA templates

The first generation DNA templates were produced through the insertion of the SA-cDNA-pA structures into plasmids that contained the HAs of each locus (Fig. 12).

The plasmids with HAs were custom-designed and purchased from *Eurofins Scientific* (*Luxembourg City, Luxembourg*). Two plasmids were acquired, one for each HAs-insert. The insert for *Tymp* contained 860 nucleotide-long HAs; while the insert for *Alb* contained 840 nucleotide-long HAs. The *Tymp* plasmid was called pEX-K4-Tymp and contained a kanamycin resistance (*kanR*) marker; while the *Alb* plasmid was called pEX-A2-Alb and contained an ampicillin resistance (*ampR*) marker. In both cases, the right and left HAs were separated by two restriction enzyme targets: SpeI and AgeI. Moreover, the whole HAs-insert was flanked by another two restriction enzyme targets (NotI and MluI), which allowed the excision of the whole HAs-insert from the plasmid backbone (Fig. 12).

The SA-cDNA-pA constructs were obtained through overlapping polymerase chain reaction (PCR). Then those constructs were cloned between the HAs through the SpeI and AgeI restriction targets (detailed on section 4.1.2.4) (Fig. 12). Two different cDNAs were used (*eGFP* and h*TYMP*). The *eGFP* cDNA was amplified from pHAGE-PGK-GFP-IRES-LUC-W plasmid (*Addgene #46793*). The h*TYMP* cDNA was amplified from p305-TP [313], a plasmid kindly provided by Dr. R. Martí (*Institut de Recerca de Vall d'Hebron (VHIR), Barcelona, Spain*). The pA SV40 element was amplified from pCMV/Bsd plasmid (*V51020, Invitrogen*). A total of four DNA templates were generated:

- ***Tymp* locus** (pEX-K4-Tymp): pEX-TYMP-GFP (*eGFP*) and pEX-TYMP-TYMP (h*TYMP*).
- ***Alb* locus** (pEX-A2-Alb): pEX-ALB-GFP (*eGFP*) and pEX-ALB-TYMP (h*TYMP*).

These first generation DNA templates could only be inserted by HDR. They had long HAs and contained the complete cDNA sequence (with their translation initiation codon ATG). In the case of *Tymp* locus, the correct insertion of the template through HDR would result in a hybrid mRNA with only one ATG codon. This mRNA would have exon 1 from genomic *Tymp* and then the complete cDNA. Since the endogenous *Tymp* ATG is located at exon 2, the hybrid mRNA would only have the ATG from the cDNA. Consequently, the resulting protein would be exclusively encoded by the cDNA (Fig. 13). In the case of *Alb* locus, if the DNA template was correctly introduced through HDR, a hybrid mRNA with two ATGs would be generated. This mRNA would have exon 1 from *Alb* and then the complete cDNA. In this case two different proteins could be synthesized. The templates were designed so that cDNAs would stay in-frame with the initial *Alb* ATG, so both alternatives would be viable. Since the first exon from *Alb* encodes a secretory peptide that is later excised, the final protein obtained would be mainly the one encoded in the cDNA (Fig. 14).

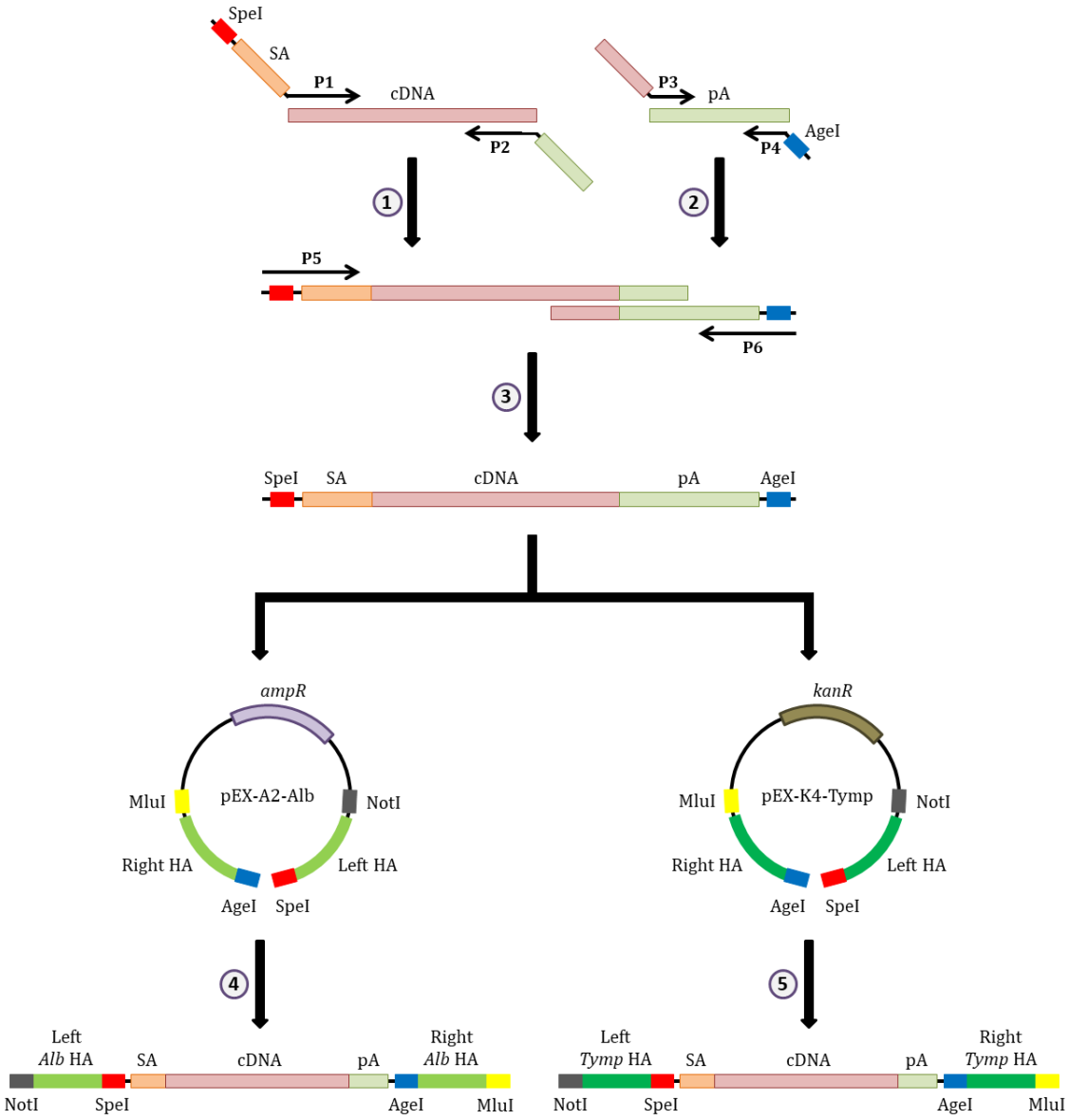

Figure 12. Production of the 1st generation DNA templates. The first step was the production of two simple PCR products with overlapping extremes. 1) The first product was the SA-cDNA element. The forward (F) primer (P1) of this reaction contained the SpeI restriction target and the SA sequence; while the reverse (R) primer (P2) contained a small sequence homologous to the pA element. 2) The second product was the pA element. The F primer (P3) of this reaction contained a small sequence homologous to the SA-cDNA element; while the R primer (P4) contained the AgeI restriction target. 3) By performing an overlapping PCR with additional primers (P5 and P6), the SA-cDNA-pA constructs flanked by SpeI and AgeI were obtained. Finally, these constructs were cloned between the HAs of the custom-plasmids: 4) pEX-A2 (*Alb* HAs); or 5) pEX-K4 (*Tymp* HAs). The sequences of all elements of the DNA templates were specified in Table S2. The restriction target sequences were specified in Table S3. The primer sequences were specified in Table 9 and 11.

4.1.2.1.2. Second generation DNA templates

The second generation DNA templates incorporated some optimizations. First, only *eGFP* cDNA was used, which simplified the manufacturing process. These DNA templates were meant to be simple tools to assess gene editing efficiency *in vitro*. If they demonstrated good results, additional templates with *TYMP* cDNA would be generated. Second, the plasmid backbone was changed to avoid background fluorescence. Theoretically, cDNAs contained in the DNA templates could only be expressed when they were correctly inserted in the desired locus. Inside the plasmids, the cDNAs would remain silenced, but when they were correctly inserted, the endogenous promoters of *Tymp* and *Alb* should transcribe them. It was expected that after the transfection of the CRISPR/Cas9 RNAs with the DNA templates, only the correctly edited cells would become fluorescent. Surprisingly, we observed that the transfection of the plasmid templates alone resulted in a high percentage of fluorescent cells. This was attributed to "leaky" expression of *eGFP* cDNA in the plasmid template. It was a big problem, as the fluorescence was not necessarily indicating the presence of edited cells. In theory, the pEX plasmid backbones did not include any promoter, but due to this leaky expression, we decided that another backbone would be more suitable. Third, these DNA templates incorporated some elements that made them susceptible to be inserted by NHEJ. The first generation templates could only be inserted by HDR, as they were delivered to cells as circular molecules. However, these new plasmids incorporated the corresponding CRISPR/Cas9 target at the 5'-end of the template sequence. This way, a cell that was transfected with both the CRISPR/Cas9 RNAs and the DNA template would be able to linearize the plasmid, so the lineal template could be introduced by NHEJ. Fourth, the HAs were optimized to be functional in case the DNA template was introduced by NHEJ. The first generation DNA templates had long HAs that encompassed the circumventing exons of intron 1 (exons 1 and 2). This was not a problem if the templates were introduced by HDR, but these new templates were engineered to be able to be introduced by HDR or NHEJ indistinctively. If the same HAs were used in these second generation templates, their insertion by NHEJ would result in an aberrant coding sequence. Exon 1 and 2 would be partially duplicated, which could pose a problem during RNA splicing or protein translation. This problem was tackled by different approaches for the two target loci. For *Tymp* locus, the whole HAs were deleted. This way, this template would only be able to be introduced by NHEJ. Since non-dividing cells are more likely to repair DSB by NHEJ, this strategy was expected to be feasible (Fig. 13). For *Alb* locus, the HAs were shortened to only encompass the sequence of intron 1. This way, if the template was introduced by NHEJ only the intron would be partly duplicated, which would not a problem as the duplicated sequence would be excised during RNA splicing (Fig. 14). In both loci, if the DNA templates were introduced correctly, the hybrid mRNAs encoded in the edited genes would be the same as the ones generated by the first generation templates.

Taking into account the mentioned optimizations, two new DNA template plasmids were designed: pC-TYMP-GFP and pC-ALB-GFP. They were generated through the insertion of the DNA template constructs into a plasmid backbone called pC, which comes from the pCMV/Bsd plasmid (*V51020, Invitrogen*). The new templates were obtained through PCR with specific primers using some first generation plasmids as templates. Then these

constructs were cloned into pC backbone through the XhoI and EcoRI restriction targets (detailed on section 4.1.2.4) (Fig. 15).

The pCMV/Bsd plasmid (*V51020, Invitrogen*) was a commercial product whose backbone was used to generate the second generation DNA plasmids. The backbone of this plasmid (pC) contained an Ori and the *ampR* gene. The insert of this plasmid contained the cytomegalovirus (CMV) promoter, the EM7 promoter, the blasticidin resistance (*bsdR*) gene and the SV40 pA. The insert was flanked by two multi-cloning sites (MCSs) that contained several restriction targets. The XhoI and EcoRI targets were used to substitute the whole insert by the desired DNA templates (Fig. 15).

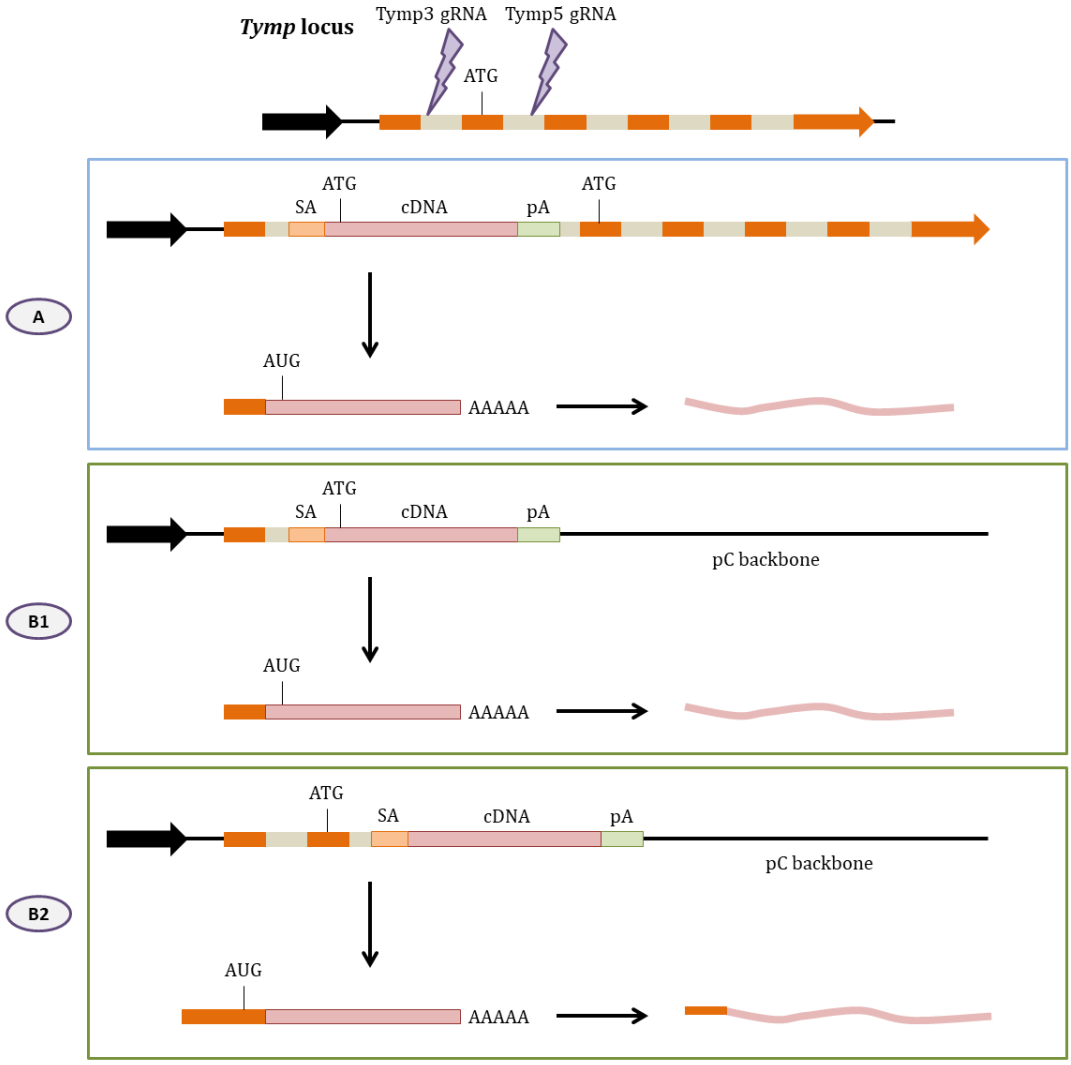

Figure 13. Genome editing of the *Tymp* locus with 1st, 2nd and 3rd generation DNA templates: scheme of the resulting genomic loci and the mRNAs and proteins generated. The DNA templates could be integrated by either HDR (A) or NHEJ (B). The 1st generation DNA templates could only be introduced through HDR (A), while the ones from the 2nd (B1) and 3rd (B2) generation could only be integrated trough NHEJ. In all cases a hybrid mRNA would be generated, but only with 3rd generation DNA templates the resulting protein would be a hybrid containing the second genomic *Tymp* exon (B2).

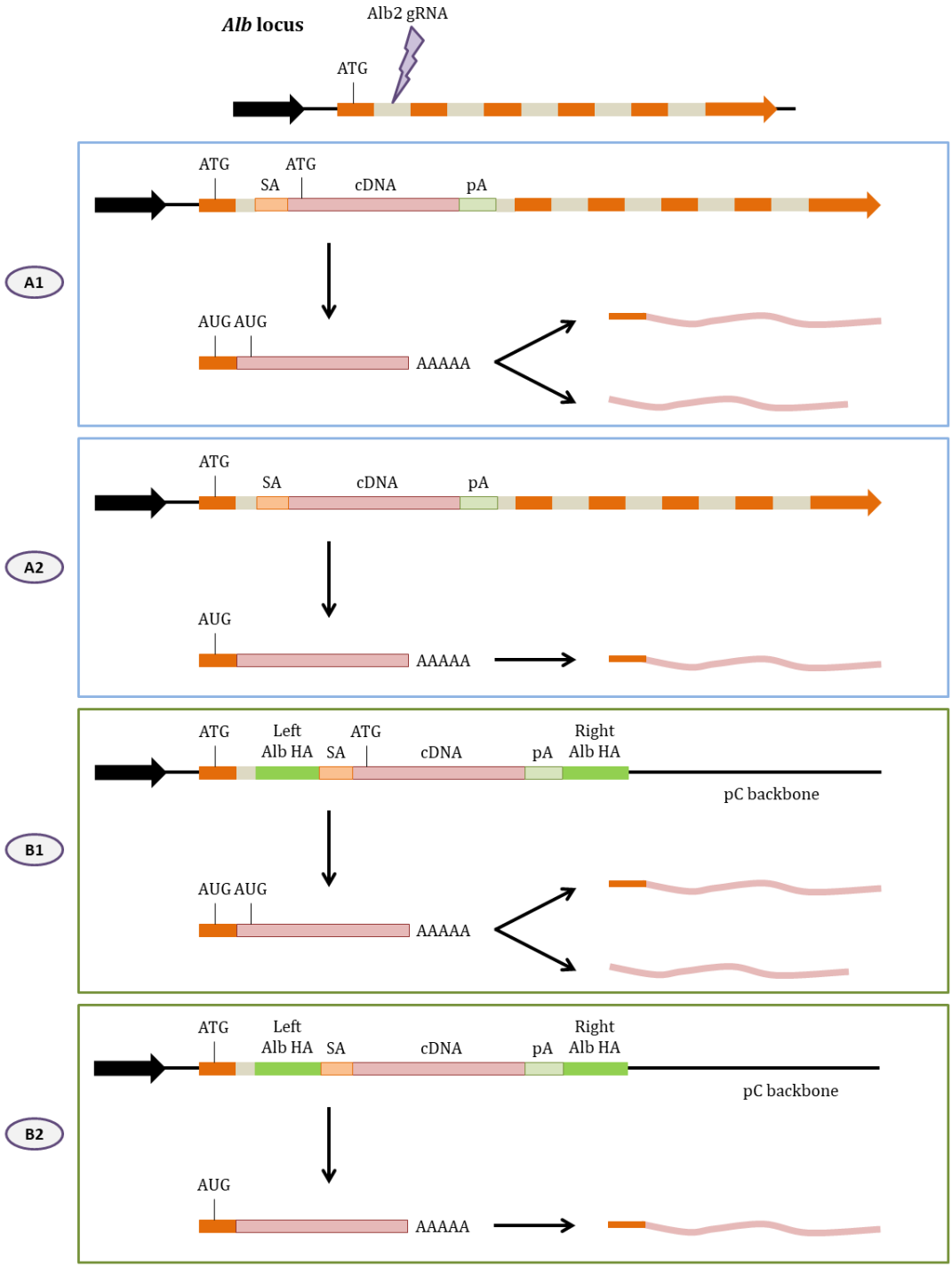

Figure 14. Genome editing of the *Alb* locus with 1st, 2nd and 3rd generation DNA templates: scheme of the resulting genomic loci and the mRNAs and proteins generated. The DNA templates could be integrated by either HDR (A) or NHEJ (B). The 1st generation templates could only be introduced through HDR (A1). The 2nd generation templates could either be integrated through HDR (A1) or NHEJ (B1). The 3rd generation templates could either be introduced through HDR (A2) or NHEJ (B2). In all cases a hybrid mRNA would be generated. A hybrid protein containing the first *Alb* exon could be generated in all cases. But with 1st and 2nd generation DNA templates, if the second ATG was used, it would be possible to synthesize a normal protein.

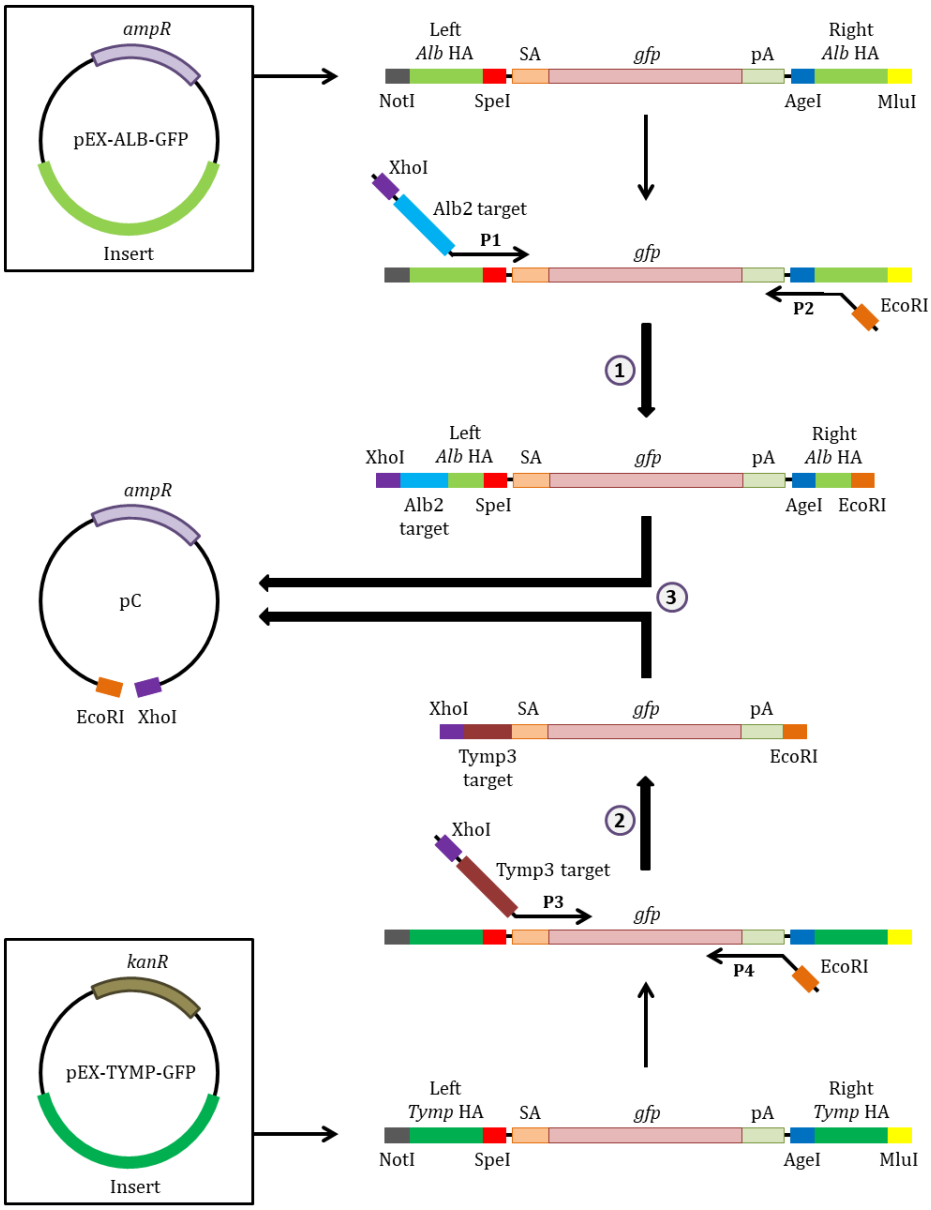

Figure 15. Production of the 2nd generation DNA templates. 1) The insert of pC-ALB-GFP was generated with a simple PCR. The F primer (P1) of this reaction contained the XhoI restriction target and the Alb2 gRNA target sequence; while the R primer (P2) contained the EcoRI restriction target. 2) The insert of pC-TYMP-GFP was generated with a simple PCR. The F primer (P3) of this reaction contained the XhoI restriction target and the Tymp3 gRNA target sequence; while the R primer (P4) contained the EcoRI restriction target. 3) Both inserts were cloned into a pC backbone, and the plasmids pC-ALB-GFP and pC-TYMP-GFP were generated. The sequences of all elements of the DNA templates were specified in Table S2. The restriction target sequences were specified in Table S3. The gRNA target sequences were specified in Table 8. The primer sequences were specified in Table 9.

4.1.2.1.3. Third generation DNA templates

The third generation DNA templates incorporated more optimizations. The change of plasmid backbone did not reduce the fluorescence background, so we tested another strategy. We hypothesized that the elimination of the endogenous initiation codon (ATG) of *eGFP* would heavily reduce the background. To this end, we performed a directed mutagenesis and the ATG codon was eliminated. Second, the gRNA of the *Tymp* locus was modified to adapt to the new situation with the deleted ATG codon. The previous gRNA (Tymp3) targeted intron 1. The endogenous ATG of *Tymp* is located at exon 2, which means that it would come after the inserted cDNA. In this locus the translation in edited cells always started with the ATG of the cDNA. Since this codon was eliminated to reduce background fluorescence, the editing approach had to be changed. For this reason, gRNAs targeting intron 2 of *Tymp* were designed and tested, and a new gRNA was selected (Tymp5). With this new strategy, the cDNA would be introduced after exon 2, which is where the initiation codon is located. The hybrid mRNA generated with this strategy would be different from the previous one. In this case a hybrid protein would be formed with a small Tymp peptide (coded in genomic exon 2) followed by the peptide encoded by the cDNA (Fig. 13). On the other hand, there was no need to change the editing strategy of *Alb* locus, as exon 1 already had an initiation codon (Fig. 14). Third, two ATGs present in the left HA of the *Alb* DNA template were removed by directed mutagenesis. To decrease even more the chances of *eGFP* expression from the plasmid template, we performed an analysis of in-frame ATGs. In the *Alb* DNA template two in-frame ATGs were located inside the left HA. We decided that their elimination was advantageous as it did not have a deleterious effect on the gene editing approach.

Taking into account the mentioned optimizations, two new DNA template plasmids were designed: pM-TYMP-GFP and pM-ALB-GFP. They were generated with different methodologies.

The pM-TYMP-GFP plasmid was generated through the insertion of the DNA template construct into the pC plasmid backbone. The new template was obtained through PCR with specific primers using the pM-ALB-GFP plasmid as template. Then the construct was cloned into pC backbone through the XhoI and EcoRI restriction targets (detailed on section 4.1.2.4) (Fig. 16). The pM-ALB-GFP plasmid was generated through the modification of the pC-ALB-GFP. Two rounds of directed mutagenesis were needed. In the first round the intrinsic ATG of *eGFP* cDNA was mutated, and in the second round the two in-frame ATGs of the left HA were changed (detailed on section 4.1.2.5) (Fig. 16).

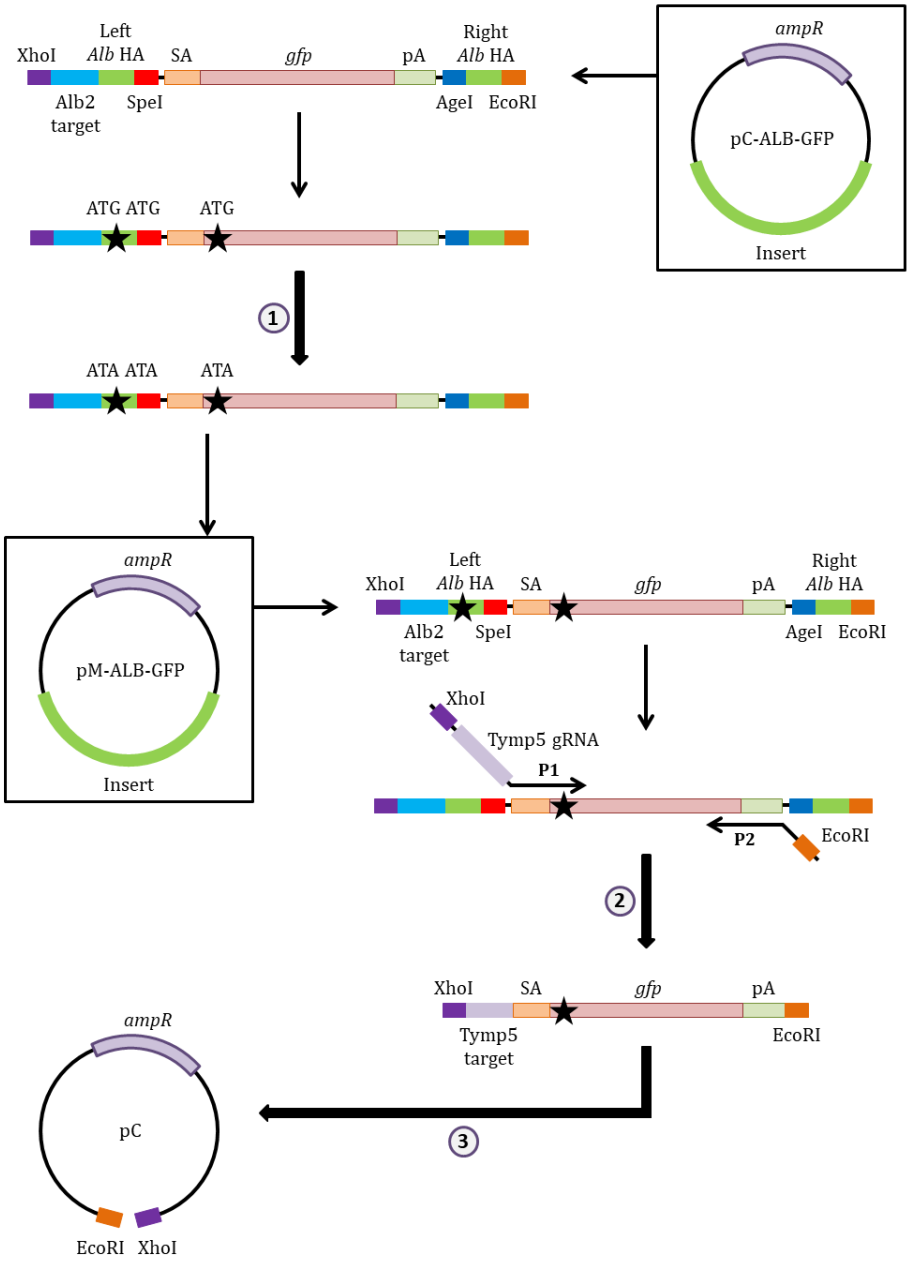

Figure 16. Production of the 3rd generation DNA templates. 1) The plasmid pM-ALB-GFP was generated through two rounds of directed mutagenesis. Three ATG codons were transformed to ATA codons. 2) This plasmid was used as a template in a simple PCR to generate the pM-TYMP-GFP insert. The F primer (P1) of this reaction contained the XhoI restriction target and the Tymp5 gRNA target sequence; while the R primer (P2) contained the EcoRI restriction target. 3) This insert was cloned into a pC backbone and pM-TYMP-GFP was generated. The sequences of all elements of the DNA templates were specified in Table S2. The restriction target sequences were specified in Table S3. The gRNA target sequences were specified in Table 8. The primer sequences of the directed mutagenesis were specified in Table 14. The primer sequences of the simple PCR were specified in Table 9. ATA: isoleucine codon.

4.1.2.1.4. Fourth generation DNA templates

Taking into account all the previous optimizations, the DNA templates meant for *in vivo* experiments were generated. These fourth generation templates presented two major changes. First, a second gRNA target was added at the 3'-end of the DNA template. This was mainly useful during *in vitro* experiments, as the whole template construct was intended to be linearized and inserted by NHEJ without the plasmid backbone. Since AAV vectors would deliver the DNA templates as lineal molecules, the presence of gRNA targets at both ends would not be necessary to linearize the templates in the *in vivo* experiments. However, their presence allowed the excision of flanking ITRs, which could be advantageous. If DNA templates without flanking gRNA targets were introduced by NHEJ, the ITRs would also be introduced. But with the presence of the gRNA targets, it would be possible to integrate the template without ITRs. For this reason, we decided to keep these structures for the *in vivo* experiments. Second, the new DNA templates contained the *TYMP* cDNA instead of *eGFP*. Two different strategies were used for each locus. For *Tymp* locus, a partial m*Tymp* cDNA (exons 3-11) was used. Although TP has the same function in humans and mice, their cDNA sequence is quite different. We decided that using the mouse sequence of *Tymp* would be advantageous for two reasons. First, by using the murine cDNA, the possibility of repairing the gene by HDR was opened. The *Tymp* DNA templates did not contain HAs, so they previously could only be integrated by NHEJ. But by using m*Tymp* cDNA, the integration by HDR was also possible, as the exons of the edited locus would be highly homologous to the introduced cDNA. Second, the use of a partial m*Tymp* cDNA allowed a neat gene editing, as the cDNA introduced did not contain the sequences of exons 1 and 2. If the gene editing was successful and the template was introduced inside intron 2, the resulting mRNA would have the endogenous murine sequence (exons 1-2 from the genomic DNA, and exons 3-11 from the template introduced). Consequently, the resulting TP formed would not be a hybrid (Fig. 17). For *Alb* locus, a cropped h*TYMP* cDNA was used. This cDNA did not contain the first 15 nucleotides (5 codons), and was designed to stay in-frame with the *Alb* exon 1. This deletion was done to eliminate the intrinsic initiation codon and to better adapt to the putative hybrid protein that would be formed. TP is initially synthesized as a polypeptide with a length of 482 amino acids. The first 10 amino acids correspond to a pro-peptide that is excised during protein maturation. For this reason, the abolition of the first 15 nucleotides was expected to have no deleterious effect in TP biosynthesis or activity. If the gene editing was successful, a hybrid mRNA composed by the first *Alb* exon and the resected h*TYMP* cDNA would be formed. The hybrid protein synthesized (Alb-hTP) would contain: Alb secretory peptide, partial Alb pro-peptide, a partial TP pro-peptide, and the complete TP polypeptide (Fig. 18).

With these changes in mind, two new DNA template plasmids were designed: pM-TYMP-TYMP and pM-ALB-TYMP. They were generated with different methodologies.

The generation of the pM-TYMP-TYMP plasmid required the construction of an intermediate plasmid (pINT-1). The first step of the procedure was the generation of the SA-m*Tymp*-pA construct, which was obtained through overlapping PCR. The m*Tymp* cDNA was amplified from the pCMV3-mTymp-HA plasmid (*MG50923-CY, Sino Biological*). The pA SV40 element was amplified from pCMV/Bsd plasmid (*V51020, Invitrogen*). Then this construct was cloned into the pC backbone through the XhoI and EcoRI restriction targets.

The resulting plasmid was pINT-1. An additional step was necessary to introduce the flanking gRNA targets. Using pINT-1 as template, the pM-TYMP-TYMP construct was generated with a simple PCR with specific primers. Finally, the construct was cloned into the pC backbone through the XhoI and EcoRI restriction targets (detailed on section 4.1.2.4) (Fig. 19). The generation of the pM-ALB-TYMP plasmid also required the construction of an intermediate plasmid (pINT-2). The first step of the procedure was the amplification by simple PCR of the cropped h*TYMP* cDNA with specific primers using pEX-ALB-TYMP plasmid as template. This construct was cloned into the backbone of pM-ALB-GFP through the SpeI and AgeI restriction targets. The resulting plasmid was pINT-2. This intermediate step allowed the introduction of the cropped h*TYMP* cDNA between the HAs that have the two mutated in-frame ATGs. Using pINT-2 as template, the pM-ALB-TYMP construct was generated with a simple PCR with specific primers. Finally, the construct was cloned into the pC backbone through the XhoI and EcoRI restriction targets (detailed on section 4.1.2.4) (Fig. 20).

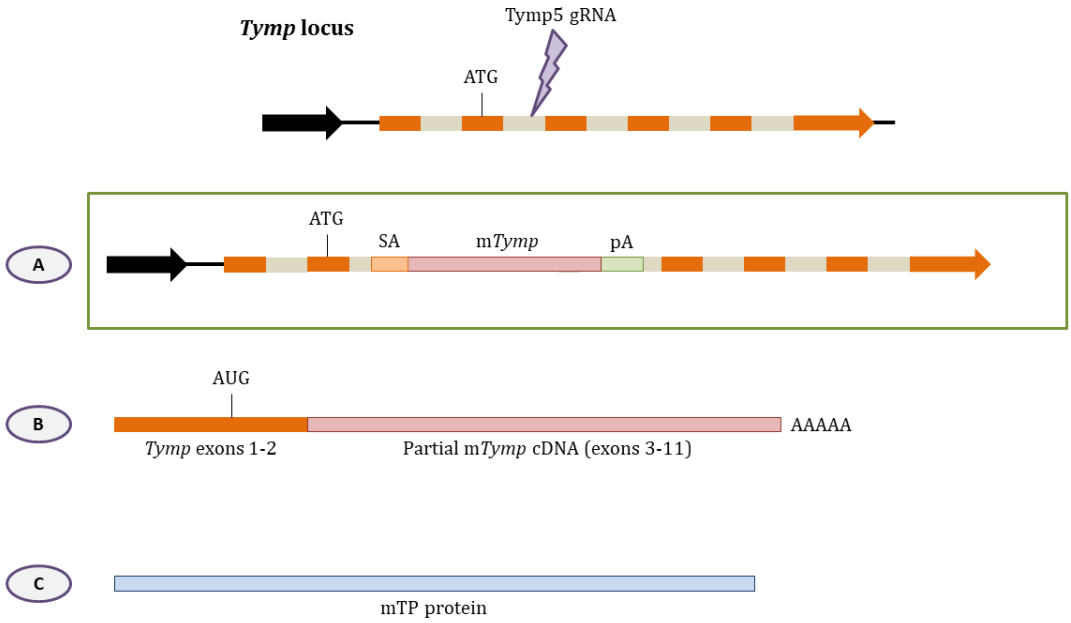

Figure 17. Genome editing of the *Tymp* locus with pM-TYMP-TYMP template. A) Edited *Tymp* locus: pM-TYMP-TYMP template would be integrated by NHEJ, as it did not contain HAs. But since the cDNA of the template is homologous to *Tymp* exons, it would also be possible to be integrated by HDR. B) Hybrid mRNA: The resulting mRNA would be composed of genomic *Tymp* exons 1-2 and the partial m*Tymp* cDNA introduced (exons 3-11). C) TP protein: The resulting protein would have the same sequence as the native TP. mTP: murine TP.

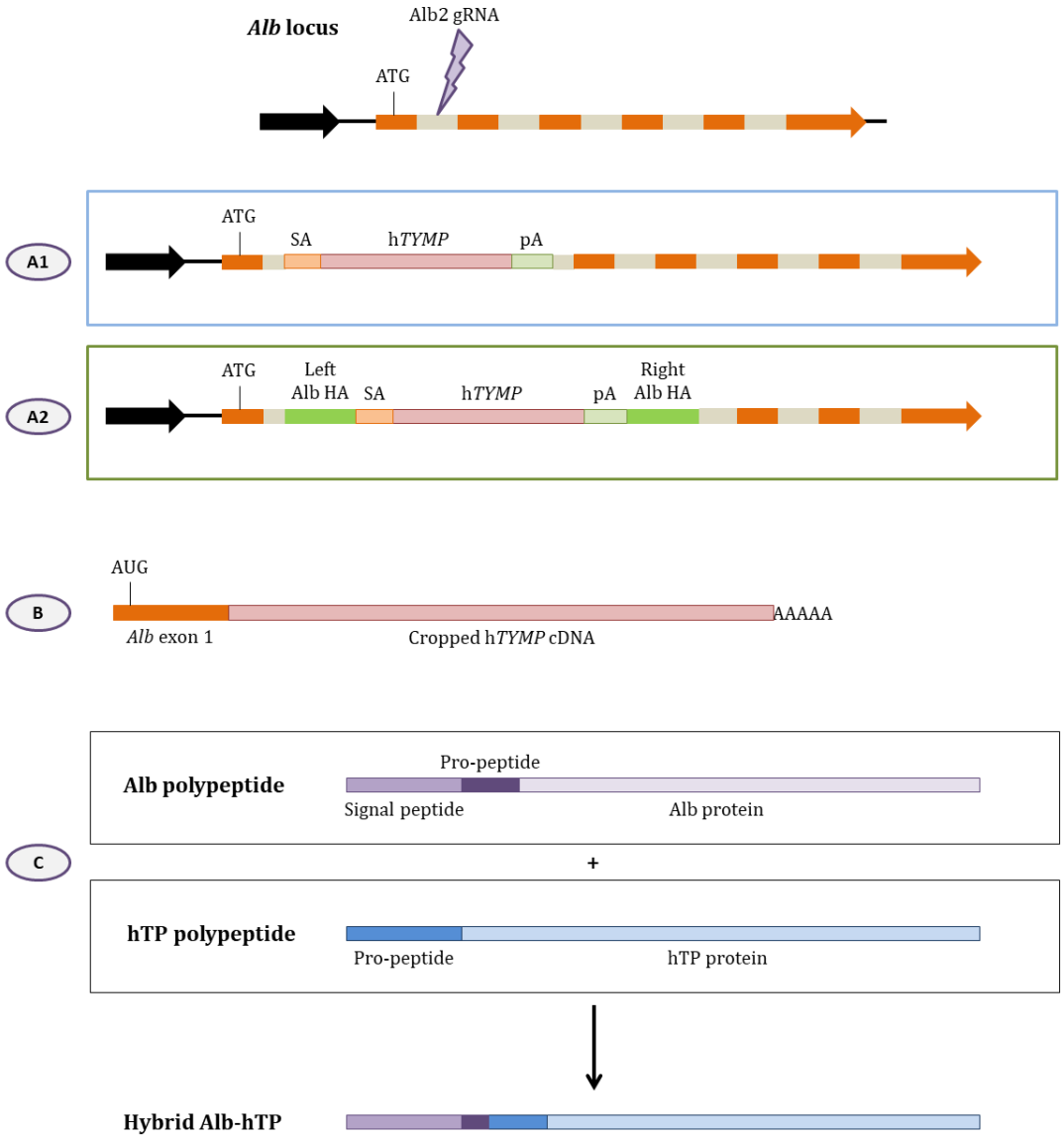

Figure 18. Genome editing of the *Alb* locus with pM-ALB-TYMP template. A) Edited *Alb* locus: pM-ALB-TYMP template would be integrated by either HDR (A1) or NHEJ (A2). B) Hybrid mRNA: With both integration methods, the same hybrid mRNA would be generated. This mRNA would be composed of *Alb* exon 1 and the cropped h*TYMP* cDNA introduced. C) Hybrid Alb-hTP protein: The resulting protein would contain a small part of Alb protein (signal peptide and fraction of the pro-peptide) and part of hTP (fraction of the pro-peptide and the functional hTP protein). hTP: human TP.

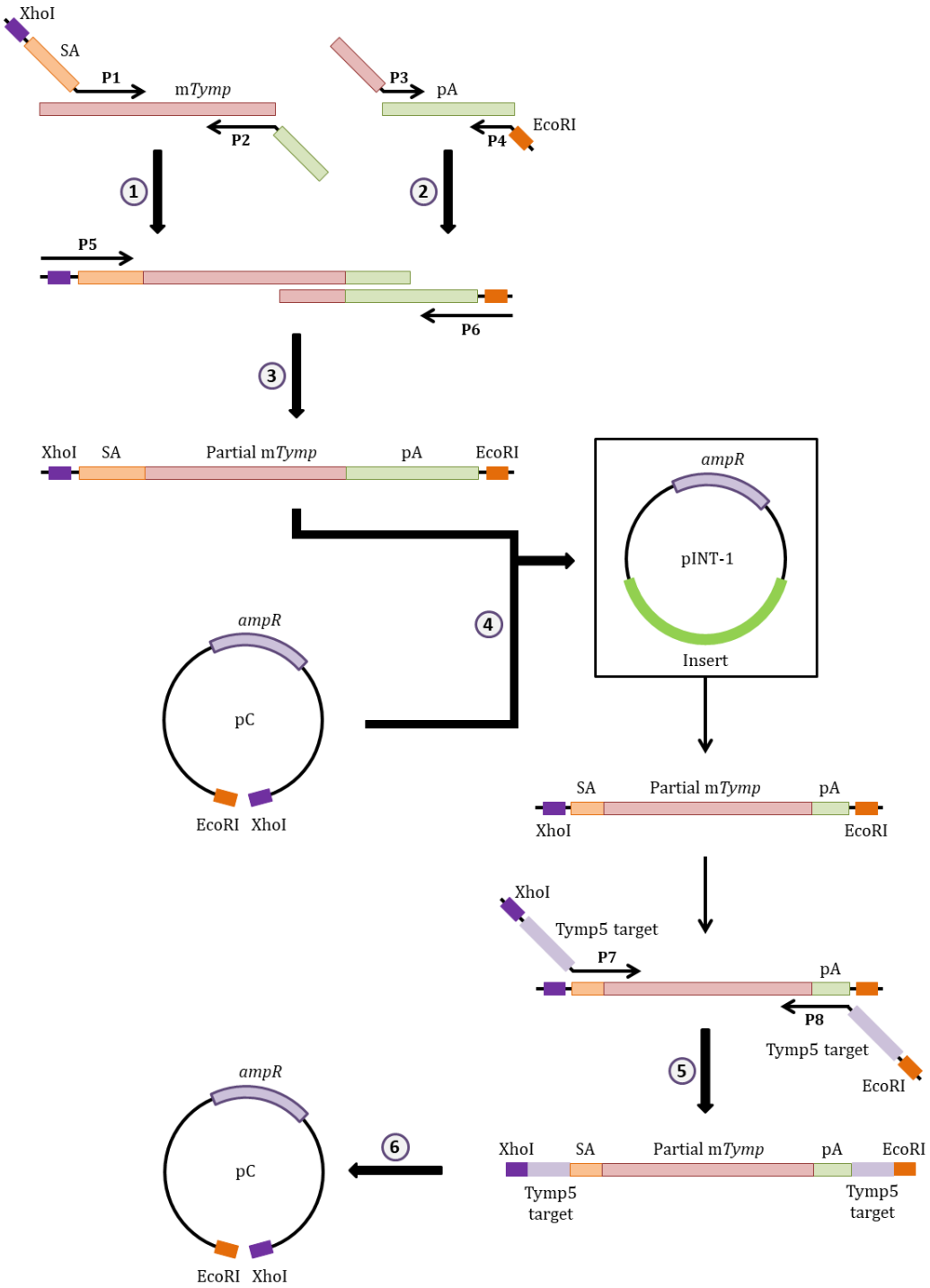

Figure 19. Production of the pM-TYMP-TYMP DNA template. 1) The insert of pINT-1 was generated with an overlapping PCR. The first step was the production of two simple PCR products with overlapping extremes. 1) The first product was the SA-m*Tymp* element. The F primer (P1) of this reaction contained the XhoI restriction target and the SA sequence; while the R primer (P2) contained a small sequence homologous to the pA element. 2) The second product was the pA element. The F primer (P3) of this reaction contained a small sequence homologous to the SA-m*Tymp* element; while the R primer (P4) contained the EcoRI restriction target. 3) By performing an overlapping PCR with additional primers (P5 and P6), the SA-m*Tymp*-pA constructs flanked by XhoI and EcoRI were obtained. 4) The pINT-1 plasmid was generated with the cloning of this construct into the pC backbone. 5) The insert of pM-TYMP-TYMP was generated with a simple PCR

using pINT-1 as template. The F primer (P7) of this reaction contained the XhoI restriction target and the Tymp5 gRNA target sequence; while the R primer (P8) contained the EcoRI restriction target and the Tymp5 gRNA target sequence. 6) The pM-TYMP-TYMP plasmid was generated with the cloning of this insert into the pC backbone. The sequences of all elements of the DNA templates were specified in Table S2. The restriction target sequences were specified in Table S3. The gRNA target sequence was specified in Table 8. The primer sequences were specified in Table 9 and 11.

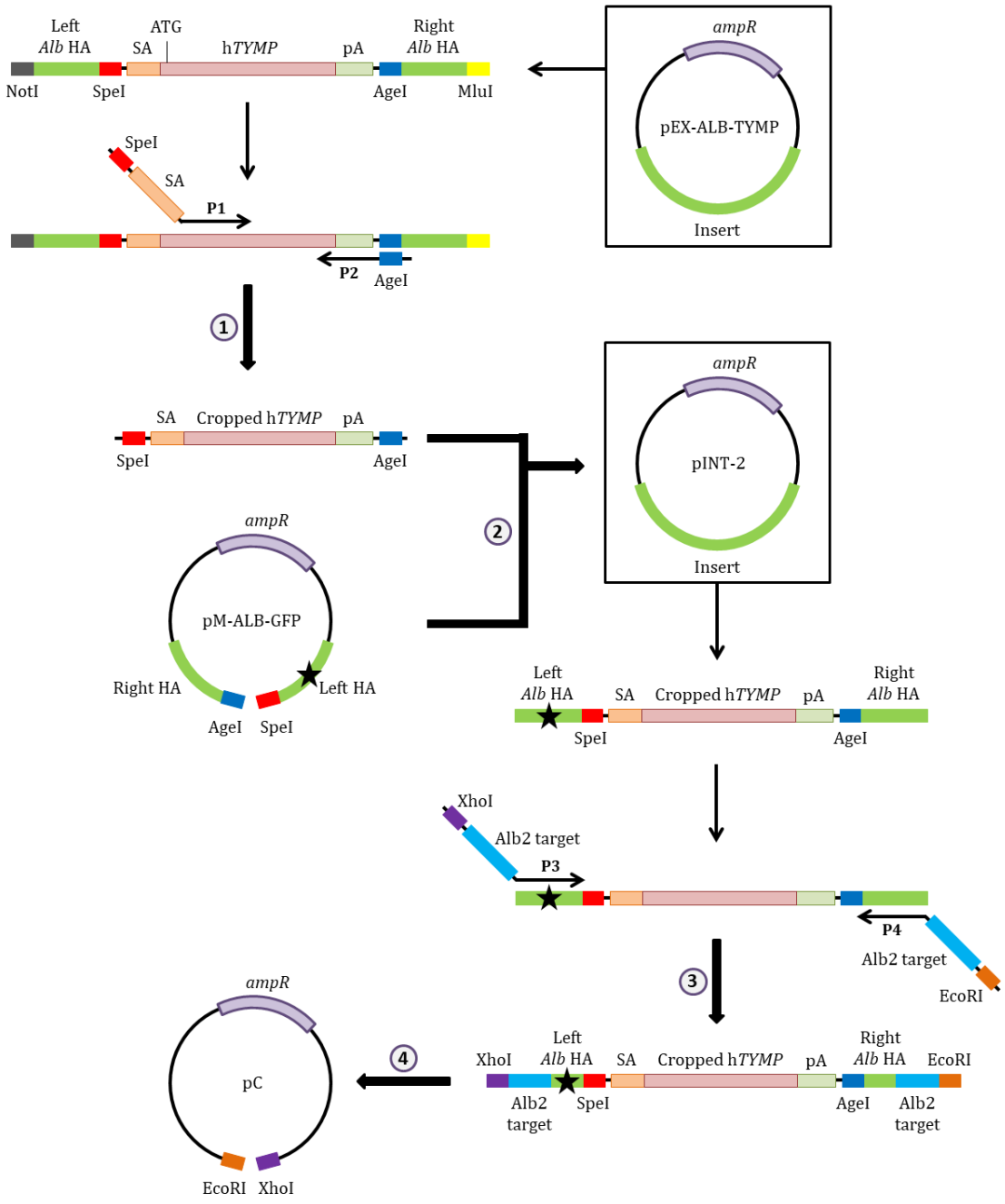

Figure 20. Production of the pM-ALB-TYMP DNA template. 1) The insert of pINT-2 was generated with a simple PCR using pEX-ALB-TYMP as template. The F primer (P1) of this reaction contained the SpeI restriction target and the SA sequence; while the R primer (P2) contained the AgeI restriction target. 2) The pINT-2 plasmid was generated with the cloning of this insert into the pM-ALB-GFP backbone, between the HAs. 3) The insert of pM-ALB-TYMP was generated with a simple PCR using pINT-2 as template. The F primer (P3) of this reaction contained the XhoI

restriction target and the Alb2 gRNA target sequence; while the R primer (P4) contained the EcoRI restriction target and the Alb2 gRNA target sequence. 4) The pM-ALB-TYMP plasmid was generated with the cloning of this insert into the pC backbone. The sequences of all elements of the DNA templates were specified in Table S2. The restriction target sequences were specified in Table S3. The gRNA target sequence was specified in Table 8. The primer sequences were specified in Table 9. Star: two in-frame initiation codons mutated (ATG → ATA).

4.1.2.2. Design of pCMV-Alb-TP control plasmid

Since the gene editing approach for the *Alb* locus implied the generation of a hybrid Alb-hTP protein, it was necessary to assess the viability of this hybrid before *in vivo* experimentation. The approach for *Tymp* locus did not imply the synthesis of a hybrid protein, so it was not necessary to generate a control plasmid.

To assess the functionality of the hybrid Alb-hTP enzyme it was necessary to recreate the resulting sequence of the gene editing. The insert of pCMV-Alb-TP control plasmid was the *Alb* exon 1 and partial intron 1, followed by the sequence of the pM-ALB-TYMP template. This construct represented the resulting sequence of a successful gene editing by HDR. The insert was introduced downstream a CMV promoter, so the coding sequence of the Alb-hTP hybrid was under the control of a strong promoter (Fig. 21). This way, it would be easy to test if the hybrid protein was functional *in vitro*.

To generate the pCMV-Alb-TP plasmid the first step was the generation of the insert by overlapping PCR. The resulting construct was composed of *Alb* exon 1 and partial intron 1 followed by the SA-h*TYMP*-pA element. *Alb* exon 1 and partial intron 1 were amplified from genomic DNA, while the SA-h*TYMP*-pA element was amplified using pM-ALB-TYMP as template. The backbone used derived from the pCMV/Bsd plasmid (*V51020, Invitrogen*). In this case, apart from the previous pC backbone elements (*ampR* and Ori), the CMV promoter was also maintained in the structure. This new backbone was called pCMV. Since the preferred restriction targets were not present at the desired locations, the pCMV backbone was also amplified by PCR and the targets were added within the primers sequence. Finally, the insert was cloned into the pCMV backbone through the MluI and AgeI restriction targets (Fig. 21).

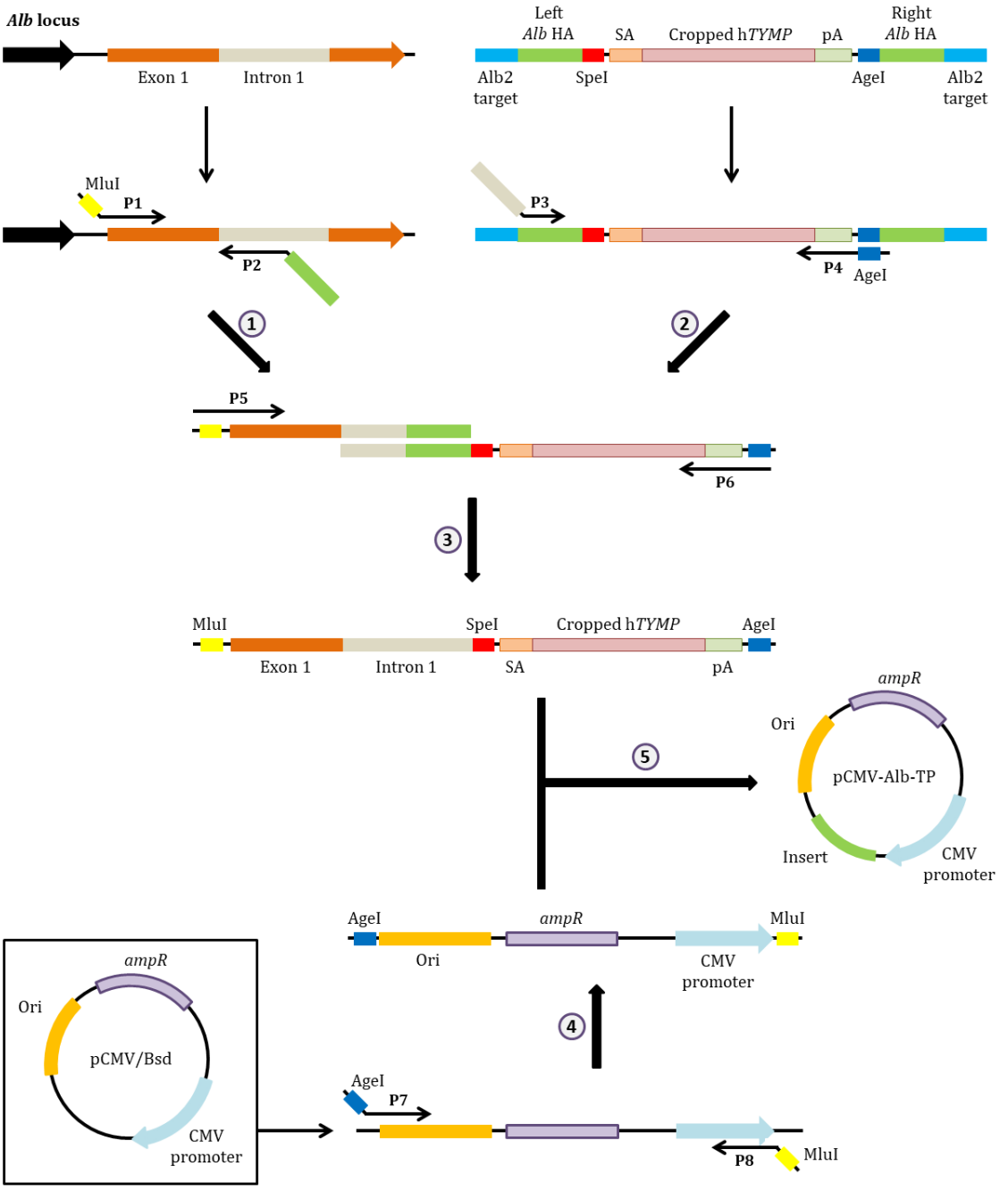

Figure 21. Production of the pCMV-Alb-TP control plasmid. The insert of pCMV-Alb-TP was generated by overlapping PCR. The first step was the production of two simple PCR products with overlapping extremes. 1) The first product was the exon 1 and partial intron 1 of *Alb* locus. The F primer (P1) of this reaction contained the MluI restriction target; while the R primer (P2) contained a small sequence homologous to the SA-h*TYMP*-pA element. 2) The second product was the SA-h*TYMP*-pA element. The F primer (P3) of this reaction contained a small sequence homologous to the first PCR product; while the R primer (P4) contained the AgeI restriction target. 3) By performing an overlapping PCR with additional primers (P5 and P6), the Alb-hTP construct flanked by MluI and AgeI were obtained. 4) The pCMV backbone was generated with a simple PCR using pCMV/Bsd as template. The F primer (P7) of this reaction contained the AgeI restriction target; while the R primer (P8) contained the MluI restriction target. 5) The pCMV-Alb-TP plasmid was generated with the ligation of the Alb-hTP construct and the pCMV backbone. The restriction target sequences were specified in Table S3. The primer sequences were specified in Table 9 and 11.

4.1.2.3. Design of plasmids for dual rAAV2/8 approach

The first gene editing approach consisted in the co-administration of rAAV2/8 vectors carrying the DNA templates with NPs carrying the CRISPR/Cas9 elements. But later, we also proposed a dual rAAV2/8 vector approach. In this case, two vectors carrying the DNA templates and the coding sequences of Cas9 and gRNA scaffolds would be administered. It was necessary to split these elements into two rAAV2/8 vectors due to the large size of the *cas9* gene. The sequence of an engineered Cas9 with optimized efficacy was used (SpCas9-HF1).

A total of 3 vectors were custom-designed and ordered as synthetic plasmids: pAAV-Cas9, pAAV-TYMP5 and pAAV-ALB2. All these constructs were flanked by ITRs, which allowed their direct use to produce rAAV2/8 vectors without additional cloning steps (Fig. 22, Table S4). These plasmids were purchased from *ATG:biosynthetics* (*Merzhausen, Germany*).

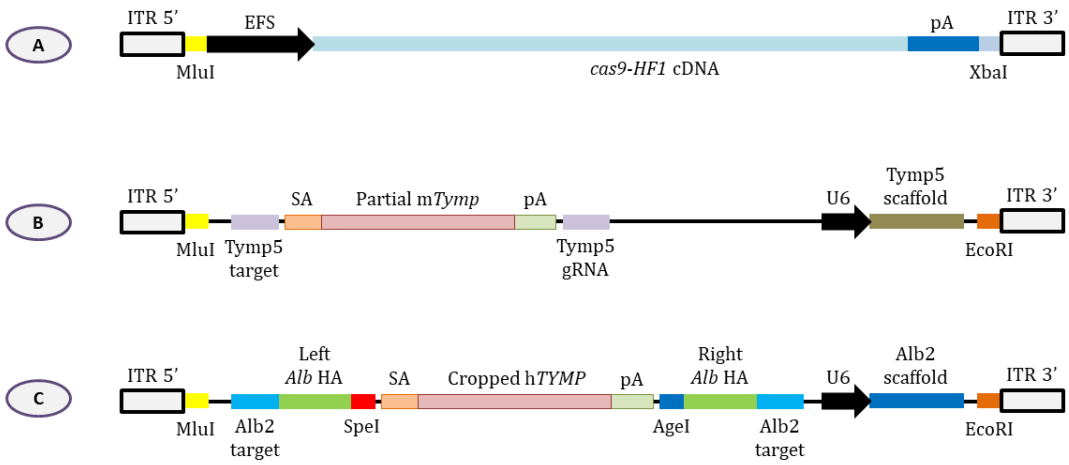

Figure 22. Schematic structure of the plasmids used in the dual AAV approach. A) pAAV-Cas9: contained the elongation factor 1α short (EFS) promoter, the *cas9-HF1* cDNA and the pA of β-globin. B) pAAV-TYMP5: contained the DNA template for *Tymp* locus (from pM-TYMP-TYMP template plasmid), the U6 promoter and the Tymp5 gRNA scaffold. C) pAAV-ALB2: contained the DNA template for *Alb* locus (from pM-ALB-TYMP template plasmid), the U6 promoter and the Alb2 gRNA scaffold. The sequences of all elements of the AAV vectors were specified in Table S2 and S4. The restriction target sequences were specified in Table S3. The gRNA target sequences were specified in Table 8.

4.1.2.4. Cloning procedures

Most of the plasmids mentioned above were generated through cloning procedures. First, the desired insert and/or plasmid backbone were amplified by PCR. The primers used were specific for each application, and they often carried additional sequences like restriction enzyme targets. Sometimes, an additional step of overlapping PCR was needed to obtain the final insert. These PCR products were purified, and then the enzymatic restriction of both insert and backbone plasmid was performed. Next, the digested fragments were run through electrophoresis, and the fragments of the desired size were

extracted from the gel and purified. Then, the backbone and the insert were ligated and transformed into *E. coli*. The bacteria were cultured in a Petri dish and colonies were selected. Each colony was a clone that contained a unique ligated plasmid. Several clones were further amplified and purified. Finally, all plasmids were sequenced to assess which of them had the correct sequence.

In the following sections this procedure is explained with more detail, including the specific details of each plasmid that was generated.

4.1.2.4.1. PCR

PCR is a widely used technique that allows the rapid multiplication of the number of copies of a specific DNA sequence. This method uses DNA polymerases and custom-designed primers to amplify DNA fragments. All PCRs were performed with *Phusion DNA Polymerase* (*F530, Thermo Scientific*), following the manufacturer's instructions. All primers were purchased from *Thermo Fisher Scientific* (*Waltham, USA*). The apparatus used to perform all reactions was the *Veriti 96-Well Thermal Cycler* (*Applied Biosystems, Foster City, USA*).

If both the desired backbone and insert were already present in distinct plasmids and flanked by the correct restriction targets, the PCR steps could be avoided.

4.1.2.4.1.1. Simple PCR

The PCR conditions were the same for all reactions except the annealing temperature, which depends on the primers used; and the extension time, which depends on the size of the product (Fig. 23, Table 9).

More often than not, the insert or the plasmid backbone didn't have the desired restriction enzyme targets flanking them, so there was a need to add them through PCR. The addition of these targets on primers was useful to obtain the constructs flanked by restriction targets. Sometimes the primers were also used to introduce small additional sequences (Table 10).

4.1.2.4.1.2. Overlapping PCR

If the desired insert was present in a single plasmid and the only problem was that it was not flanked by the correct restriction targets, a simple PCR would suffice to solve the problem. But in some cases it was necessary to generate an insert that contained sequences found in different sources. In these cases, the problem could be solved by performing an overlapping PCR. With this method it was possible to couple two different PCR products. The only requisite was to design the primers so that the two initial fragments had homologous extremes.

The overlapping PCR procedure required two distinct PCR programs. In the first-round PCR the conditions were the same for all reactions except the extension time (Fig. 24, Table 11). In the second-round PCR the conditions were the same for all reactions except the annealing temperature and the extension time (Fig. 23, Table 11). During the first-round PCR (Fig. 24), the two single PCR products were added in the same reaction at equimolar concentration without primers. The fragments acted as both templates and primers to each other, and a unified DNA fragment was obtained. When the first PCR program ended, specific primers were added to the reaction (Table 12) and a second-round PCR was performed. This second PCR was carried out with the same program as the simple PCR (Fig. 23). While the first PCR program goal was to couple both fragments, the second PCR round goal was to further amplify the final DNA product.

Plasmids	Constructs	Product size (bp)	Annealing temperature (ºC)	Extension time (s)
DNA templates (1st generation)	SA-*eGFP*	784	60,4	25
	SA-h*TYMP*	1516	69	45
	pA	163-164	52,6	10
DNA templates (2nd generation)	pC-TYMP-GFP	922	56	25
	pC-ALB-GFP	1572	45	45
DNA templates (3rd generation)	pM-TYMP-GFP	922	59,1	30
DNA templates (4th generation)	SA-m*Tymp*	1295	65,4	40
	pA	156	55,8	5
	m*Tymp* + gRNAs	1459	51	45
	h*TYMP* alone	1626	52,9	50
	h*TYMP* + HAs	2312	47,8	70
pCMV-Alb-TP control plasmid	*Alb* exon 1 and partial intron 1	490	57	15
	SA-h*TYMP*-pA	1674	56	50
	pCMV backbone	2573	64,6	78

Table 9. Simple PCR program characteristics for each plasmid generation.

Plasmids	Constructs	Primer structure (5'-3')	
DNA templates (1st generation)	SA-*eGFP*	F	GTAG**ACTAGT**TTCTCTCCCTGTTTCCACAGCGGCCGCCATGGTGAGCAAG SpeI SA
		R	AAGCTGCAATAAACAAGTAGGGATCCTTACTTGTACAG Homology with pA

	SA-h*TYMP*	F	GGTA**ACTAGT**TTCTCTCCCTGTTTCCACAGCGATGGCAGCCTTGATGACC SpeI SA
		R	AAGCTGCAATAAACAAGTTTCGCGGCAAAGGAGCTTTA Homology with pA
	pA (*eGFP*)	F	CTGTACAAGTAAGGATCCCTACTTGTTTATTGCAGCTT Homology with *eGFP*
		R	CCTAAG**ACCGGT**ACAGACATGATAAGATACATTG AgeI
	pA (*TYMP*)	F	TAAAGCTCCTTTGCCGCGAAACTTGTTTATTGCAGCTT Homology with *TYMP*
		R	GGCTAG**ACCGGT**ACAGACATGATAAGATACATTG AgeI
DNA templates (2nd generation)	pC-TYMP-GFP	F	CTGA**CTCGAG**ATAGGTACTAGACGTGGGCGGGGTTCTCTCCCTGTTTCCAC XhoI Tymp3 gRNA target
		R	GCACT**GAATTC**ACAGACATGATAAGATACATTGAT EcoRI
	pC-ALB-GFP	F	CTGA**CTCGAG**CCTGCTCGACCATGCTATACTAAGTATTTTTCTAGTAATGG XhoI Alb2 gRNA target
		R	GCACT**GAATTC**ACCACACAACATATTTAA EcoRI
DNA templates (3rd generation)	pM-TYMP-GFP	F	CTGA**CTCGAG**CCCATACGCACCTCAGGGATCCATTCTCTCCCTGTTTCCAC XhoI Tymp5 gRNA target
		R	GCACT**GAATTC**ACAGACATGATAAGATACATTGAT EcoRI
DNA templates (4th generation)	SA-m*Tymp*	F	GGTA**CTCGAG**TTCTCTCCCTGTTTCCACAGGGGCCATGCTAATGGCCATT XhoI SA
		R	TAAGCTGCAATAAACAAGTCGAGACTTAGGGCTGTGCG Homology with pA
	pA (m*Tymp*)	F	CGCACAGCCCTAAGTCTCGACTTGTTTATTGCAGCTTA Homology with mTymp
		R	GCTAG**GAATTC**ACAGACATGATAAGATACATTG EcoRI
	m*Tymp* + gRNAs	F	CTGA**CTCGAG**CCCATACGCACCTCAGGGATCCATTCTCTCCCTGTTTCCAC XhoI Tymp5 gRNA target
		R	GCAC**GAATTC**TGGATCCCTGAGGTGCGTATGGGACAGACATGATAAGATAC EcoRI Tymp5 gRNA target
	h*TYMP* alone	F	GGTA**ACTAGT**TTCTCTCCCTGTTTCCACAGTGACCCCGGGAACCGGGGCC SpeI SA
		R	GGCTAG**ACCGGT**ACAGAC AgeI
	h*TYMP* + HAs	F	CTGA**CTCGAG**CCTGCTCGACCATGCTATACTAAGTATTTTTCTAGTAATGG XhoI Alb2 gRNA target
		R	AGCT**GAATTC**TTAGTATAGCATGGTCGAGCAGGACCACACAACATATTTAA EcoRI Alb2 gRNA target
pCMV-Alb-TP control plasmid	*Alb* exon 1 and partial intron 1	F	GCTAG**ACGCGT**GATCTACAGTTATTGGTTAAAGA MluI
		R	GGCCCTATGAGACCGTAATA *Alb* intron 1
	SA-h*TYMP*-pA	F	GACTCAAATTACGTTGGAT *Alb* intron 1
		R	TAGTATAG**ACCGGT**ACAGAC AgeI
	pCMV backbone	F	GCTAT**ACCGGT**GACGAAAGGGCCTCGTGATA AgeI
		R	GCGAT**ACGCGT**ACGGTCCGTTCCAATGCACC MluI

Table 10. Simple PCR primer characteristics for each plasmid generation.

Plasmids	Constructs	Product size (bp)	Annealing temperature (ºC)	Extension time (s)
DNA templates (1st generation)	SA-*eGFP*-pA	909	55,4	25
	SA-h*TYMP*-pA	1679	57,6	45
DNA templates (4th generation)	SA-m*Tymp*-pA	1413	59	45
pCMV-Alb-TP control plasmid	Alb-hTP coding sequence	2114	58	64

Table 11. Overlapping PCR program characteristics for each plasmid generation.

Plasmids	Constructs	Primer structure (5'-3')	
DNA templates (1st generation)	SA-*eGFP*-pA	F	GTAG**ACTAGT**TTCTCTCCCTGT SpeI SA
		R	CCTAAG**ACCGGT**ACAGACAT AgeI
	SA-h*TYMP*-pA	F	GGTA**ACTAGT**TTCTCTCCCTGT SpeI SA
		R	GGCTAG**ACCGGT**ACAGAC AgeI
DNA templates (4th generation)	SA-m*Tymp*-pA	F	GGTA**CTCGAG**TTCTCTCCCT XhoI SA
		R	GCTAG**GAATTC**ACAGACATG EcoRI
pCMV-Alb-TP control plasmid	Alb-hTP coding sequence	F	GCTAG**ACGCGT**GATCTACAG MluI
		R	TAGTATAG**ACCGGT**ACAGAC AgeI

Table 12. Overlapping PCR primer characteristics for each plasmid generation.

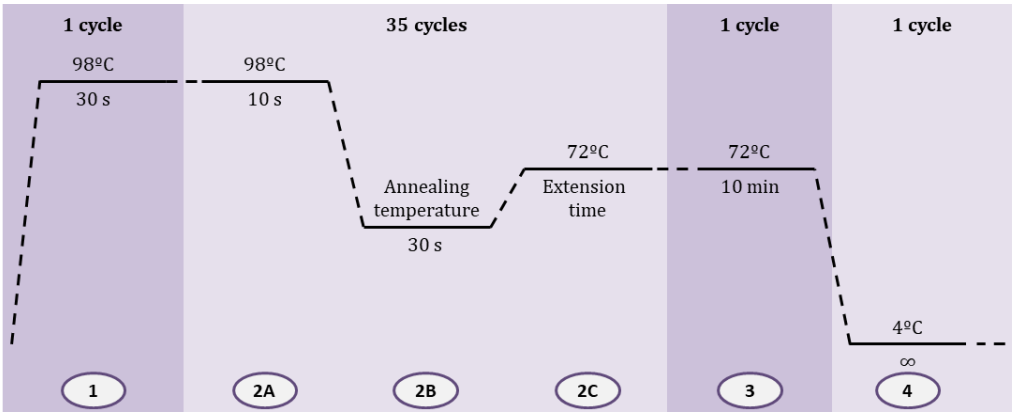

Figure 23. Simple PCR program. 1) Initial denaturalization; 2a) Denaturalization; 2b) Primer annealing: the annealing temperatures of the different PCR reactions were specified in Table 9 (simple PCR) and Table 11 (second-round overlapping PCR); 2c) Extension: the extension times of the different PCR reactions were specified in Table 9 (simple PCR) and Table 11 (second-round overlapping PCR); 3) Final extension; 4) Stand-by. ∞: infinite time.

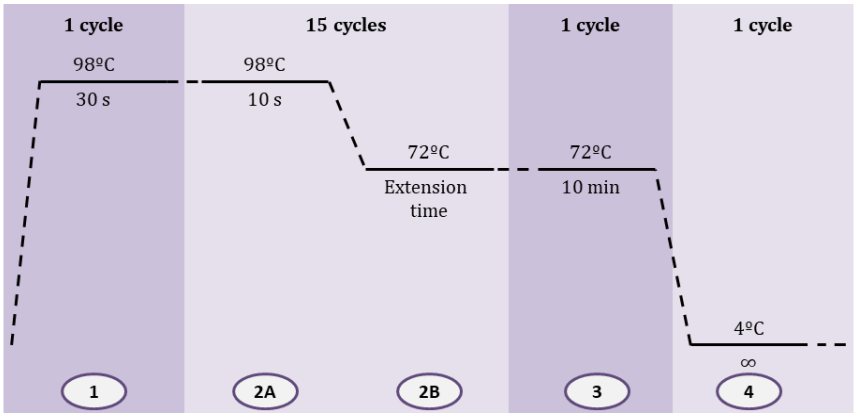

Figure 24. First-round overlapping PCR program. 1) Initial denaturalization; 2a) Denaturalization; 2b) Annealing and extension: the extension times of the different PCR reactions were specified in Table 11; 3) Final extension; 4) Stand-by.

4.1.2.4.1.3. Verification and purification of the PCR products

The confirmation that PCR reactions yielded the correct DNA products was always needed before proceeding to the next steps. A simple method to ascertain this was by running a small proportion of the PCR product in an agarose gel electrophoresis (detailed on section 4.7.6). The presence of a single band with the expected size confirmed that the PCR reaction was successful. The absence of a band, or the presence of multiple non-specific bands, determined that the PCR reaction had to be repeated with modified conditions. When a successful PCR was verified, the PCR products were purified to remove the reagents used in the PCR reactions (detailed on section 4.7.7.2).

4.1.2.4.2. Enzymatic restriction

When the correct plasmid backbone and insert were obtained, the next step was their enzymatic restriction. All the plasmid generation approaches required a double digestion with two different restriction enzymes. This strategy allowed the ligation of the insert in a unique orientation. There are several restriction enzymes available in the market, and also several restriction buffers. Each enzyme has preference for a different buffer. In the case of double digestion, it was necessary to find the most suitable buffer so that both enzymes had an optimal activity. The tool *Double Digest Calculator* developed by *Thermo Scientific* (*Waltham, USA*) was used to determine the best enzymatic restriction conditions for each application (Table 13). The restriction reactions were incubated at 37ºC overnight. The next day the restriction enzymes were inactivated by incubating the reactions at 80ºC for 20 minutes.

Plasmids	Backbone	Insert	Restriction enzymes		Buffer
DNA templates (1st generation)	pEX-K4	SA-*eGFP*-pA	SpeI (*ER1252, Thermo Scientific*)	AgeI (*ER1461, Thermo Scientific*)	Tango
		SA-h*TYMP*-pA			
	pEX-A2	SA-*eGFP*-pA			
		SA-h*TYMP*-pA			
DNA templates (2nd generation)	pC	pC-TYMP-GFP	XhoI (ER0692, *Thermo Scientific*)	EcoRI (*1040, Takara Bio*)	Tango
		pC-ALB-GFP			
DNA templates (3rd generation)	pC	pM-TYMP-GFP	XhoI (ER0692, *Thermo Scientific*)	EcoRI (*1040, Takara Bio*)	Tango
DNA templates (4th generation)	pC	SA-m*Tymp*-pA	XhoI (ER0692, *Thermo Scientific*)	EcoRI (*1040, Takara Bio*)	Tango
		m*Tymp* + gRNAs			
		h*TYMP* + HAs			
	pM-ALB-GFP	h*TYM*P alone	SpeI (*ER1252, Thermo Scientific*)	AgeI (*ER1461, Thermo Scientific*)	Tango
pCMV-Alb-TP control plasmid	pCMV	Alb-hTP coding sequence	MluI (*ER0561, Thermo Scientific*)	AgeI (*ER1461, Thermo Scientific*)	R

Table 13. Enzymatic restriction characteristics for each plasmid generation. All the buffers mentioned were provided with the restriction enzymes.

4.1.2.4.3. Dephosphorylation of the plasmid backbone

After enzymatic restriction, the plasmid backbones were dephosphorylated with shrimp alkaline phosphatase (SAP) (*783901000UN, Applied Biosystems*), following the manufacturer's instructions. The reactions were incubated at 37ºC for 1 hour, and then at 65ºC for 20 minutes to inactivate SAP. This step was necessary to avoid the plasmid backbone self-ligation.

4.1.2.4.4. DNA purification by gel electrophoresis

The restricted plasmid backbones and inserts were run in an agarose gel electrophoresis (detailed on section 4.7.6). Then the respective desired bands were excised and the DNA fragments were purified (detailed on section 4.7.7.1).

4.1.2.4.5. Ligation

Each pair of plasmid backbone and insert was ligated with a T4 DNA ligase (*15224090, Invitrogen*), following the manufacturer's instructions. These ligation reactions were performed in a 1:5 molar ratio [backbone : insert]. They were incubated at 16ºC overnight (Table 14).

4.1.2.4.6. *E. coli* chemical transformation and clonal selection

The ligated plasmids obtained in the previous section were introduced to *5-alpha Competent E. coli* (*C2987I, New England Biolabs (NEB)*) trough chemical transformation. A small proportion of each ligation reactions were mixed with *E. coli*. These mixtures were incubated on ice for 30 minutes, and then a heat shock was applied (incubation at 42ºC for 30 seconds). The tubes were further incubated on ice for 5 minutes. Next, S.O.C. medium (*15544034, Invitrogen*) was added, and the mixtures were incubated at 37ºC for 1 hour with a shaking speed of 250 rpm (revolutions per minute). Finally, the transformed *E. coli* were cultured in a Petri dish and incubated at 37ºC overnight. The Petri dishes contained a mixture of lysogeny broth (LB) medium and *Select Agar* (*30391023, Invitrogen*). This solid bacterial growth medium also contained the appropriate selection antibiotic. This medium allowed the replication of single *E. coli* cells that had introduced a molecule of ligated plasmid. The next day several discrete colonies were visible, and each one of them corresponded to the progeny of an initial *E. coli* cell that had successfully incorporated a particular ligated molecule, also referred as plasmid clone. Since not all clones were correct, several of them had to be selected to be sure that at least one of them had the correct ligated plasmid. For this reason, 3-8 colonies were selected and mixed with liquid LB broth that contained the appropriate selection antibiotic. All clones were incubated at 37ºC overnight with a shaking speed of 250 rpm.

LB medium was manufactured as follows. To obtain 1 L of LB broth, the following reagents had to be mixed in 1 L of water: 10 g of tryptone (*48647, Serva*), 5 g of yeast extract (*Y1625, Sigma-Aldrich*), and 10 g of sodium chloride (NaCl) (*S3014, Sigma-Aldrich*). To manufacture LB-Agar medium, 15 g of *Select Agar* (*30391023, Invitrogen*) was added to 1 L of LB broth. Those bacterial growth media contained the appropriate antibiotic for each application. Ampicillin (Amp) (*A9518, Sigma-Aldrich*) was added at a concentration of 100 µg/ml; while kanamycin (Kan) (*K1377, Sigma-Aldrich*) was added at a concentration of 50 µg/ml.

Plasmids	Backbone	Insert	Resulting plasmid
DNA templates (1st generation)	pEX-K4	SA-*eGFP*-pA	pEX-TYMP-GFP
		SA-h*TYMP*-pA	pEX-TYMP-TYMP
	pEX-A2	SA-*eGFP*-pA	pEX-ALB-GFP
		SA-h*TYMP*-pA	pEX-ALB-TYMP
DNA templates (2nd generation)	pC	pC-TYMP-GFP	pC-TYMP-GFP
		pC-ALB-GFP	pC-ALB-GFP
DNA templates (3rd generation)	pC	pM-TYMP-GFP	pM-TYMP-GFP
DNA templates (4th generation)	pC	SA-m*Tymp*-pA	pINT-1
		m*Tymp* + gRNAs	pM-TYMP-TYMP
		h*TYMP* + HAs	pM-ALB-TYMP
	pM-ALB-GFP	h*TYM*P alone	pINT-2
pCMV-Alb-TP control plasmid	pCMV	Alb-hTP coding sequence	pCMV-Alb-TP

Table 14. Ligation characteristics for each plasmid generation.

4.1.2.4.7. Plasmid DNA purification and clonal verification

The plasmids contained in each *E. coli* clone were purified with the *mi-M-Plasmid Miniprep Kit* (*mi-PMN50, Metabion*), following the manufacturer's instructions. This kit provided a simple method for *E. coli* lysis and plasmid DNA purification using silica-based membrane columns. The plasmids were separated from the bacterial residues.

Each purified plasmid clone was sequenced (detailed on section 4.7.8.1) with the appropriate primers (Table S1): pEX-TYMP templates with primers Seq-5,6,7,8,9,10; pEX-ALB templates with primers Seq-1,2,3,4,9,10; pC-TYMP-GFP with primers Seq-11,14; pC-ALB-GFP primers Seq-11,14; pM-TYMP-GFP with primers Seq-11,14; pM-ALB-GFP with primers Seq-11,14; pINT-1 with primers Seq-11,13,14; pM-TYMP-TYMP with primers Seq-11,13,14; pINT-2 with primers Seq-11,12,14; pM-ALB-TYMP with primers Seq-11,12,14; and pCMV-Alb-TP with primers Seq-12,14. The incorrect clones were discarded.

4.1.2.5. Directed mutagenesis

When there was a need to change a small part of a DNA sequence the best choice was to perform a directed mutagenesis. This DNA modification was performed with the *GeneArt Site-Directed Mutagenesis System* (*A13282, Invitrogen*), following the manufacturer's instructions. This methodology required two complementary custom-designed primers with the desired mutation included at the centre of their sequence (Table 15). The primers were purchased from *Thermo Fisher Scientific* (*Waltham, USA*). Linearized plasmids with the mutations incorporated were obtained by performing PCRs with the specific primers. These plasmids were re-circularized through recombination and chemically transformed to *E coli*. The next steps were equivalent to the ones detailed on sections 4.1.2.4.6 and 4.1.2.4.7.

Mutation location	Original plasmid	Resulting plasmid	Primer sequence (5'-3')	
Intrinsic ATG of *eGFP* cDNA	pC-ALB-GFP	pC-ALB-GFP ATA	F	CCACAGCGGCCGCCAT**A**GTGAGCAAGGGCGAGG
			R	CCTCGCCCTTGCTCAC**T**ATGGCGGCCGCTGTGG
Two in-frame ATGs of the left *Alb* HA	pC-ALB-GFP ATA	pM-ALB-GFP	F	GGTATCTTTGAT**A**ACAATAAT**A**GGGGATTTTGAA
			R	TTCAAAATCCCC**T**ATTATTGT**T**ATCAAAGATACC

Table 15. Characteristics of the primers used for directed mutagenesis. Mutated nucleotides in bold and underlined.

4.2. NP production

4.2.1. PNPs from GEMAT-IQS

PNPs were generated at GEMAT-IQS. OM-pBAEs were synthesized following a two-step procedure, as described previously [291,292,293,294,295,296,297]. First, an acrylate-terminated polymer was synthesized by addition reaction of primary amines with diacrylates (at 1:1,2 molar ratio [amine : diacrylate]). Then, OM-pBAEs were obtained by end-capping modification of the resulting acrylate-terminated polymer with thiol-terminated oligopeptide at 1:2,1 molar ratio in dimethyl sulfoxide (DMSO). The mixture was stirred overnight at room temperature (RT) and the resulting polymer was obtained by precipitation in a mixture of diethyl ether and acetone (1:1) (Fig. 25). These OM-pBAEs were mixed with the desired RNAs and incubated at 25ºC for 30 minutes to allow the formation of electrostatic bonds. A further step of pH change allowed the precipitation of molecules in small particles. PNPs were lyophilized and stored at -20ºC in the freezer. When needed, they were resuspended in sterile water and administered to cells or animals. Initially, different combinations of OM-pBAEs were tested to determine which formulation had a better performance. During this screening step, PNPs carried the eGFP mRNA (*L-6101, Trilink Biotechnologies*). The two formulations with the best efficiency were selected. The first one, the KH PNP, consisted of two polymers with different amino acids as terminal peptide groups: lysine (K) and histidine (H). They were mixed in a 60/40 ratio [K/H polymers]. The second one, the KH-retinol (KH-ret) PNP, consisted of three different polymers. Two of them were the same as the ones of KH PNP, but the third one was a K polymer with retinol molecules (Kret). They were mixed in a 55/5/40 ratio [K/Kret/H polymers] (Fig. 25). These two PNP formulations, carrying the Cas9 mRNA and gRNAs (in either 1:2 or 1:6 weight ratios [gRNA : Cas9 mRNA]), were used for the assessment of gene editing efficiency *in vitro*. It was demonstrated that the KH-ret formulation had a better liver targeting [294], so only this PNP was used *in vivo*.

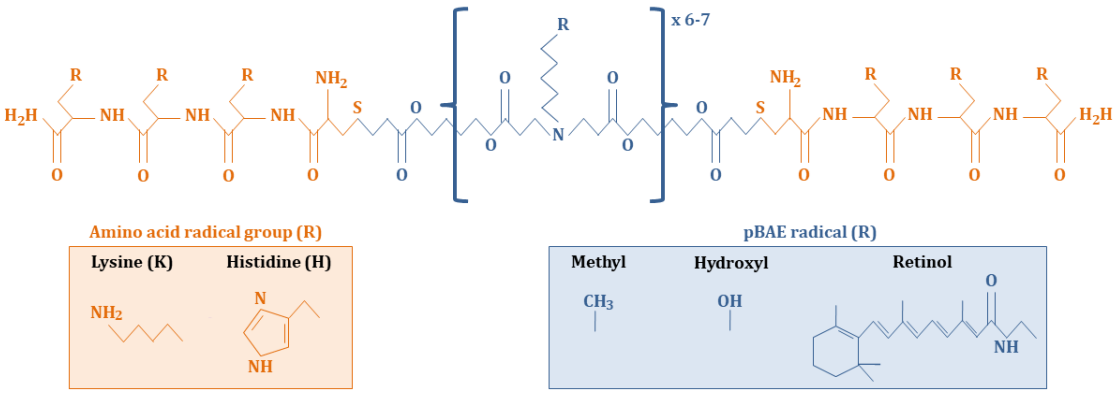

Figure 25. Structure of OM-pBAE polymers. The oligopeptide structure (orange) could have two different amino acid group radicals (**R**): K or H. In the pBAE structure (blue), the region marked with a parenthesis was repeated 6-7 times, and each repeat could have a different radical (**R**): methyl, hydroxyl or retinol. K and H polymers contained methyl and hydroxyl radicals in a 1:1 ratio, while the Kret polymer contained methyl, hydroxyl and retinol radicals in a 1:1:1 ratio [291,294]. S: sulphur.

4.2.2. LNPs from Acuitas Therapeutics®

LNPs were generated at Acuitas Therapeutics®, as described previously [306,307,308,309,310,311,312]. First, gRNAs and Cas9 mRNA were mixed in a 1:1 mass ratio in aqueous solution at pH 4. Then the RNAs were added to a combination of four lipids: cholesterol, 1,2-distearoyl-sn-glycero-3-phosphocholine (DSPC), a PEG2000 lipid with a C14 anchor, and a proprietary ionizable lipid (Fig. 26). This combination was performed in an ethanol solution at a 0,035 mRNA/lipid weight ratio by in-line mixing, allowing the combination of the lipids and RNAs to self-assemble. The resulting LNPs were subsequently dialyzed overnight in aqueous solution to remove residual ethanol. They were delivered in phosphate-buffered saline (PBS) (1x) supplemented with a cryoprotectant at a concentration of 1 µg/µl of RNA.

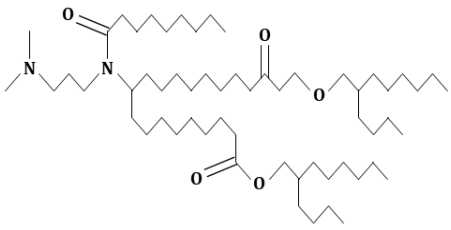

Figure 26. Structure of Acuitas Therapeutics® proprietary ionizable lipid [312].

4.3. rAAV production

All rAAV2/8 vectors were generated at Viral Vector Production Unit (UPV) located at *Centre de Biotecnologia Animal i de Teràpia Gènica* (CBATEG) – *Universitat Autònoma de Barcelona* (UAB) (*Barcelona, Spain*). The DNA constructs that were to be included in the rAAV2/8 genomes were sent to UPV as plasmids. If the constructs were not flanked by ITRs, UPV introduced them by cloning. UPV already had the AAV2/8-eGFP control vector in stock, so there was no need to send any plasmid in this case (Table 16). Vectors were produced by co-transfection of HEK-293 cells and purified by PEG precipitation and iodixanol gradient. They were delivered in PBS-MK + 40% iodixanol buffer, and stored at -80ºC in the freezer.

Application	AAV vector	Plasmid	Cloning of ITRs in UPV	Stock concentration (gc/ml)
Assessment of *in vitro* expression kinetics	AAV2/8-eGFP	-	-	$1{,}48 \times 10^{13}$
Gene editing approach 1: NPs + rAAV2/8 vectors	AAV2/8-TYMP-TYMP	pM-TYMP-TYMP	Yes	$1{,}65 \times 10^{13}$
	AAV2/8-ALB-TYMP	pM-ALB-TYMP	Yes	$1{,}43 \times 10^{13}$
Gene editing approach 2: dual rAAV2/8 vectors	AAV2/8-Cas9	pAAV-Cas9	No	$1{,}33 \times 10^{13}$
	AAV2/8-TYMP5	pAAV-TYMP5	No	$1{,}25 \times 10^{13}$
	AAV2/8-ALB2	pAAV-ALB2	No	$1{,}32 \times 10^{13}$

Table 16. Characteristics of rAAV2/8 vectors production. gc: genome copies.

4.4. Cell culture procedures

4.4.1. Cell lines and culture conditions

Cells were cultured in complete growth medium in the incubator *Heracell 150i* (*Thermo Scientific, Waltham, USA*). They were maintained at 37ºC and 5% of atmospheric carbon dioxide (CO_2) concentration. To keep cells in a sterile environment, all procedures were performed inside the biological safety cabinet *Baker SG-400 SterilGard Class II Type A/B3* (*Baker, Sanford, USA*).

Complete growth media were composed of two basic elements: the base medium and the supplements. Two types of base media were used: Dulbecco's Modified Eagle's Medium F12 (DMEM/F12) (*31331028, Gibco*), and Alpha Minimum Essential Medium (α-MEM) (*22561, Gibco*). Several supplements were added at choice: Foetal Bovine Serum (FBS) (*S181B, Biowest*), Penicillin/Streptomycin (P/S) solution (*PS-B, Capricorn Scientific*), L-glutamine (L-glu) solution (*X0550, Biowest*), ITS solution (insulin + transferrin + sodium selenite) (*I3146, Sigma-Aldrich*) and dexamethasone (*D2915, Sigma-Aldrich*).

Two different cell lines were used: Alpha Mouse Liver 12 (AML-12) (*CRL-2254, American Type Culture Collection (ATCC)*), and Hepa1c1c7 (*95090613, Sigma-Aldrich*). Each cell line was cultured in its preferred complete growth medium.

AML-12 is a murine hepatic cell line established from hepatocytes derived of a transgenic mouse (CD1 strain, line MT42) with the gene of human transforming growth factor α. These cells are adherent and have the morphology of epithelial cells. They were cultured in DMEM/F12 medium supplemented with FBS (10%), P/S (1x), ITS (1x), and dexamethasone (40 ng/µl). This cell line was used during the initial *in vitro* experiments.

Hepa1c1c7 is a murine hepatic cell line established from a hepatoma that arose in a C57/L mouse. These cells are adherent and have the morphology of epithelial cells. They were cultured in α-MEM medium supplemented with FBS (10%), P/S (1x), and L-glu (1x). Since this cell line presents a strong *Alb* gene expression, it was used in later stages of the *in vitro* experimentation to assess the efficacy of the gene editing approach.

4.4.2. Thawing and freezing

Newly purchased cell lines were received in frozen vials. The first step was the thawing and establishment of a stock. Cells could be maintained in culture during a limited time, so they had to be renewed periodically. For this reason, a newly received cell line had to be expanded and several vials had to be frozen. This way, when the limited time of cultured cells was reached, a new vial was thawed and the experimentation could continue. For long term storage, the frozen vials were kept at -196ºC inside a liquid nitrogen tank.

The thawing procedure was the following: the vial was partially submerged in a water bath at 37ºC until cells started thawing. Then the tube was transported to the biosafety cabinet and complete growth medium was added slowly, until cells were completely

thawed. Finally, cells were centrifuged at 1500 rpm for 5 minutes, the supernatant was discarded, the pellet was resuspended in fresh complete growth medium, and cells were cultured in a T75 flask.

The freezing procedure was the following: cells were recollected and centrifuged at 1500 rpm for 5 minutes. The supernatant was discarded and the pellet was resuspended in fresh complete growth medium supplemented with 10% of DMSO (*D8418, Sigma-Aldrich*). This reagent is a cryoprotectant that minimizes cell damage during freezing. These cells were introduced into different vials and stored at -80ºC in the freezer. The vials were kept inside a *Mr. Frosty Freezing Container* (*5100, Thermo Scientific*) with isopropanol (*1.09634.5000, Merck*). This recipient allowed a slow decrease of the temperature (-1 °C/min), which was beneficial for cell conservation. The next day, the vials were transported to a liquid nitrogen tank.

4.4.3. Harvest

The cell lines used in this work were adherent, so trypsinization was needed to harvest them. The process was the following: the medium was removed, and cells were washed two times with PBS (1x) (*X0515, Biowest*). This step was necessary to eliminate trypsin inhibitors present in the medium. Then PBS 1x was removed and trypsin (1x) (*X0930, Biowest*) was added. Cells were incubated for 2-5 minutes at 37ºC. During this step the trypsin removed all cell adherences. Consequently, after this incubation the cells were in suspension. Next, the action of trypsin was blocked by the addition of fresh complete growth medium. Then cells were recollected into a tube and centrifuged at 1500 rpm for 5 minutes. Finally, the supernatant was discarded and the pellet was resuspended in fresh complete growth medium.

4.4.4. Maintenance

To be able to perform experiments regularly it was necessary to keep cell lines in continuous culture. For this reason, cells were maintained in 15 ml of growth medium in T75 flasks. When they were needed for experimentation, a fraction of those cells was separated and seeded in particular multi-well plates. Cell lines cultured in the optimal conditions are continuously dividing. Consequently, their maintenance implied a frequent renewal of growth medium and periodical subculture. The frequency of medium change and subculture was determined by the growth rate of each cell line. Cells consumed the nutrients of the medium when they divided, so every two or three days it was necessary to substitute the consumed medium by fresh medium. On the other hand, adherent cells needed space to divide, so it was necessary to perform a subculture before total confluency was reached. For this reason, cells were harvested and only a fraction of cells was re-seeded in new T75 flasks. The dilution factor used in the subculture was different for each cell type. The frequency of subculture also varied between cell types, but usually it was performed twice a week. The leftover cells of the subculture were either discarded or used for experimentation.

4.4.5. Cell count and seeding

Experiments had to be reproducible, so cells were always seeded at a specific density. Since a particular cell number had to be cultured, it was necessary to quantify the harvested cells with a Neubauer chamber. First, a fraction of the harvested cells was diluted with PBS (1x). The dilution factor was empirically determined. Then this dilution was mixed with trypan blue (*T8154, Sigma-Aldrich*) at a 1:1 volume ratio. This reagent allowed the discrimination of dead cells. The final mixture was loaded into the Neubauer chamber, and cells were counted with the *TE200 Inverted Microscope* (*Nikon, Tokyo, Japan*). For accurate counting, it was necessary that the gridded squares contained between 30 and 300 cells. The average number of cells observed in the four gridded squares was annotated. The cell concentration was calculated with the following formula:

$$\text{Cell concentration (cells/ml)} = \left(\frac{\sum \text{Number of cells inside gridded square}}{4}\right) \times \text{Dilution factor} \times 10000$$

Most experiments required a 70% of cell confluency the day after seeding, so the number of cells seeded was empirically determined to attain this goal. Each type of multi-well plate and flask had a different surface area, so the number of seeded cells was adapted to each format. The type of cell line also influenced the cultured cell density (Table 17).

Cell line	Type of multi-well plate	Number of seeded cells / well	Volume of medium (µl) / well
AML-12	6 wells	450.000	2000
	12 wells	150.000	1000
	24 wells	75.000	500
	96 wells	15.000	150
Hepa1c1c7	6 wells	300.000	2000
	12 wells	100.000	1000
	24 wells	50.000	500
	96 wells	10.000	150

Table 17. Cell culture conditions for each cell line.

4.4.6. Lipofectamine transfection

For DNA transfection the reagent *Lipofectamine 3000* (*L3000015, Invitrogen*) was used, following the manufacturer's instructions. The final DNA-Lipofectamine mixture was added to the cells seeded the previous day. This procedure was used for *in vitro*

transfection of all plasmids: eGFP control plasmid (pEGFP-1), DNA templates, pCMV-Alb-TP, pAAV-Cas9, pAAV-TYMP, and pAAV-ALB.

For RNA transfection the reagent *Lipofectamine Messenger MAX* (*LMRNA150, Invitrogen*) was used, following the manufacturer's instructions. The final RNA-Lipofectamine mixture was added to the cells seeded the previous day. When it was necessary to transfect the CRISPR/Cas9 system, Cas9 mRNA and gRNAs were added together at the RNA mixture. This procedure was used for *in vitro* transfection eGFP mRNA (*L-6101, Trilink Biotechnologies*) and CRISPR/Cas9 elements (Cas9 mRNA and gRNAs).

In both cases, before the compounds were mixed, DNA/RNA molecules and Lipofectamine were diluted in *Opti-MEM* (*31985070, Gibco*), an optimized medium suitable for transfection. Moreover, a medium change was performed before transfection to minimize cell toxicity. This medium did not contain P/S, which is a mildly toxic component. Since Lipofectamine is also slightly toxic, the presence of both compounds could cause an increase of cytotoxicity. The next day the medium was changed, this time containing P/S. At that point all molecules should have been internalized, so the elimination of residual Lipofectamine reduced toxicity without hampering transfection efficiency. If the cells were to be cultured for a long time, a medium change was performed every 2-3 days.

The quantity of DNA/RNA used in a transfection procedure depended on the seeded cell density, which means it depended on the type of multi-well plate used. Moreover, each cell type had different transfection efficiency, so the optimal conditions for each cell line had to be empirically determined (Table 18). Flow cytometry was used to assess the optimal transfection conditions (reagent amounts and time of peak expression) (detailed on section 4.7.9). The pEGFP-1 plasmid was used to assess the best DNA transfection conditions, while eGFP mRNA was used to establish the best RNA transfection conditions. Using these eGFP-control molecules different setting were tested, and the ones that yielded the highest fluorescent percentage were selected for further experiments.

4.4.6.1. Transfection of the gene editing tools

The gene editing approach of NPs carrying CRISPR/Cas9 elements and rAAV2/8 vectors carrying DNA templates was simulated *in vitro* by delivering both elements with Lipofectamine. First, template plasmids were transfected with *Lipofectamine 3000*. After 18 hours, CRISPR/Cas9 elements were transfected with *Lipofectamine Messenger MAX*. Both individual procedures were performed as described in the previous section, but the reagent amounts and nucleic acid concentrations were adapted. In this case, the gene editing tools were delivered at a 1:2 molar ratio [CRISPR/Cas9 : Template], so the amounts of plasmid templates and *Lipofectamine 3000* were accordingly adjusted.

The gene editing approach of dual rAAV2/8 vectors carrying CRISPR/Cas9 elements and DNA templates was simulated *in vitro* by delivering both elements with *Lipofectamine 3000*. Since two plasmids had to be co-administered, the reagent amounts were doubled. Otherwise, the transfection procedure was the same as that described in the previous section.

Cell line	Type of multi-well plate	Type of transfection				
		DNA		RNA		
		Lipofectamine 3000 (µl)	DNA (ng)	*Lipofectamine Messenger MAX* (µl)	mRNA (ng)	gRNA (ng)
AML-12	6 wells	15	3000	15	4500	750
	12 wells	6	1200	6	1800	300
	24 wells	3	600	3	900	150
	96 wells	1	200	1	300	50
Hepa1c1c7	6 wells	7,5	2500	7,5	2500	625
	12 wells	3	1000	3	1000	250
	24 wells	1,5	500	1,5	500	125
	96 wells	0,3	100	0,3	100	25

Table 18. Transfection conditions for each cell line.

4.4.7. NP transfection

4.4.7.1. PNPs from GEMAT-IQS

To assess their transfection efficacy *in vitro*, PNPs carrying the eGFP mRNA were used in AML-12 cells. They were added to cells seeded in 96-well plates. For each well, the PNPs contained 1,3 µg of eGFP mRNA. The transfected cells were analysed by flow cytometry (detailed on section 4.7.9). These initial experiments allowed the selection of the formulations that resulted in the highest transfection efficacy, which were KH and KH-ret. Next, these two PNPs carrying the CRISPR/Cas9 elements were used. They were added to cells seeded in 24-well plates. For each well, the PNPs contained 7 µg of CRISPR/Cas9 in a 1:6 weight ratio [gRNA : Cas9].

The day after being seeded, cells were transfected with PNPs. A medium change without P/S was performed before transfection to minimize cell toxicity. The next day the medium was changed, this time containing P/S. If the cells were to be cultured for a long time, a medium change was needed every 2-3 days.

4.4.7.2. LNPs from Acuitas Therapeutics®

To assess their efficacy *in vitro*, LNPs carrying the CRISPR/Cas9 elements were used in Hepa1c1c7 cells. They were added to cells seeded in 24-well plates. For each well, the LNPs contained 7 µg of CRISPR/Cas9 at a 1:1 weight ratio [gRNA : Cas9].

The day after being seeded, cells were transfected with LNPs. A medium change without P/S was performed before transfection to minimize cell toxicity. The next day the medium was changed, this time containing P/S.

4.4.8. AAV transduction

AAV vectors were tested *in vitro* in AML-12 cells to assess their efficacy and expression kinetics. In this case, AAV2/8-eGFP vectors were used. The transduced cells were analysed by flow cytometry (detailed on section 4.7.9).

The first step was the determination of the multiplicity of infection (MOI) factor, which is the number of AAV vectors for each cell. Taking into account the desired MOI, the quantity of AAV vectors was determined using the following formula:

$$\text{Number of viral particles (gc)} = \text{MOI} \times \text{Number of seeded cells}$$

A MOI of 10^5 was used to determine eGFP expression kinetics. The desired vector quantity was diluted in growth medium supplemented with low FBS concentration (1%). AAV vectors work better in absence of FBS, but cultured cells need serum to grow. For this reason, the FBS concentration was decreased to the minimum to optimize transduction while maintaining cell viability. The medium of the cells seeded the previous day was removed and the medium with AAV vectors was added. The amount of medium used was also reduced to improve transduction. Less medium volume implied that vectors were nearer cells, which enhanced their efficiency. After 24 hours, complete growth medium with 10% FBS was added to the cells. If the cells were to be cultured for a long time, a medium change was needed each 2-3 days.

4.4.9. Blockage of protein secretion

The gene editing approach of *Alb* locus yielded a hybrid protein that contained the signal peptide of Alb. Consequently, it was expected that this protein would be secreted to the extracellular medium. This was a problem during initial experiments with *eGFP* cDNA. Since the Alb-eGFP protein was secreted, it was not possible to quantify the edited cells by fluorescence. For this reason, cells were treated with 15 µg/ml of Brefeldin A (*B7651, Sigma-Aldrich*) for 5 hours. This molecule blocked protein secretion, so the produced Alb-eGFP proteins were retained inside cells.

4.5. Animal procedures

All animal procedures were performed in accordance to the European recommendations and were approved by our institutional *Comitè Ètic d'Experimentació Animal*.

4.5.1. Mice characteristics

Mice with a C57BL/6J genetic background were used. The experimental design required the use of both WT and dKO *Tymp*$^{-/-}$ *Upp1*$^{-/-}$ [42] animals. WT mice were purchased from *Charles River Laboratories* (*Wilmington, USA*). Three dKO mice (1 male, 2 females) were kindly provided by Dr. R. Martí (*VHIR, Barcelona, Spain*). From these initial mice, a colony was generated and maintained until the end of experimentation.

4.5.2. Housing and experimentation conditions

Mice were kept in the conventional housing facility of our institution. The accommodation room had controlled humidity and temperature (22ºC) and 12-hour light-dark cycles. Each cage contained up to 5 animals. Mice were fed with the 2018 Teklad global 18% protein rodent diet, which was available *ad libitum*. Animals were segregated by sex in different cages, except the ones used to maintain the dKO colony. Both female and male mice we used in the *in vivo* experiments.

Adult mice were marked using an ear punch (*10806-290, VWR International*). The treatments were performed when they were 6-10 weeks old. The weight and physical conditions of the treated mice were monitored periodically through the entire duration of experimentation.

4.5.3. Blood sampling

The gene editing efficiency was monitored through the plasma concentration of Thd and dUrd. Blood samples were obtained from mice injected with both CRISPR/Cas9 elements and DNA templates. A sample was extracted one week before treatment. Then additional samples were extracted 2 and 4 weeks post-injection.

Mice were manually restricted and the submandibular vein was punctured with a 20-gauge (G) needle. The overflowing blood was collected by capillarity with *Microvette tubes* (*10780044, Sarstedt*). The haemorrhage was stopped by pressing the wound with a gauze pad. All samples were diluted 1:9 in PBS (1x), and then centrifuged at 3000 $_{x}$ g (times gravity) for 5 minutes at 4ºC. The plasma fraction was obtained and stored at -80ºC until analysis.

4.5.4. Intravenous injections

Mice were first immobilized with a restrainer (*MLA5016, ADinstruments*) to have easy access to the tail. Then the desired substances were injected through the lateral tail vein. Up to 200 µl were administered with a 30 G needle. Mice of the negative control group were injected with physiological saline solution (0,9% NaCl).

The lyophilized PNPs were maintained at -20ºC until use. Before administration, they were resuspended in sterile water to obtain a concentration of 250 ng/µl of RNA. Each animal was injected with 50 µg of RNA (200 µl). Two doses were tested: dose 1 required one injection, while dose 2 required two injections performed in consecutive days. Two different weight ratios [gRNA : Cas9 mRNA] were used: 1:2 and 1:6 (Table 19).

LNPs were delivered in PBS (1x) supplemented with a cryoprotectant at a concentration of 1 µg/µl of RNA. They were stored at -80ºC until use. Before administration, PNPs were diluted in saline solution (0,9% NaCl) to obtain a concentration of 250 ng/µl of RNA (200 µl). Each animal was injected with 50 µg of RNA (200 µl), using a 1:1 weight ratio [gRNA : Cas9 mRNA] (Table 19).

rAAV2/8 vectors were delivered in PBS-MK + 40% iodixanol buffer (Table 16) and stored at -80ºC until use. Before administration, vectors were diluted in saline solution to obtain the desired concentration (Table 19). Mice were injected with 200 µl of diluted vectors.

CRISPR/Cas9 efficiency was assessed with a unique injection of NPs. The first gene editing approach required the injection of rAAV2/8 vectors first, and then the injection of NPs the next day. Depending on the experimental dose, one or two NP injections were needed (on subsequent days). There was a control group injected only with rAAV2/8 vectors. The second gene editing approach required a unique injection of rAAV2/8 vectors, as both vectors were mixed in the same solution.

4.5.5. Euthanasia and tissue sampling

At the end of experimentation, mice were euthanized with $C0_2$. Right after death, necropsies were performed and the livers were surgically extracted.

For CRISPR/Cas9 efficiency assessment, livers were deposited in tubes with DMEM medium and were kept in ice. Then, they were rinsed with PBS (1x) and chopped with the help of scissors, scalpels and syringe plungers. When the tissue was properly fragmented, genomic DNA was extracted (detailed on section 4.7.1).

For gene editing experiments, each liver was fragmented and divided in three different tubes. These tubes were flash-frozen in liquid nitrogen and kept at -80 ºC until analysis. When needed, livers were thawed and processed like mentioned above. Then DNA, RNA or proteins were extracted (detailed on sections 4.7.1, 4.7.2.2 and 4.7.3.2).

Application		Dose	
		NPs	rAAV2/8 vectors
CRISPR/Cas9 efficiency	PNPs Dose 1	50 µg RNA / mouse Weight ratio 1:2 or 1:6 [gRNA : Cas9 mRNA]	-
	PNPs Dose 2	100 µg RNA / mouse Weight ratio 1:6 [gRNA : Cas9 mRNA]	-
	LNPs	50 µg RNA / mouse Weight ratio 1:1 [gRNA : Cas9 mRNA]	-
Gene editing approach 1: NPs + rAAV2/8 vectors	PNPs Dose 1	50 µg RNA / mouse Weight ratio 1:2 or 1:6 [gRNA : Cas9 mRNA]	10^{11} gc / mouse AAV2/8-TYMP-TYMP or AAV2/8-ALB-TYMP
	PNPs Dose 2	100 µg RNA / mouse Weight ratio 1:6 [gRNA : Cas9 mRNA]	2×10^{11} gc / mouse AAV2/8-TYMP-TYMP or AAV2/8-ALB-TYMP
	LNPs	50 µg RNA / mouse Weight ratio 1:1 [gRNA : Cas9 mRNA]	4×10^{11} gc / mouse AAV2/8-TYMP-TYMP or AAV2/8-ALB-TYMP
	Only rAAV2/8 vectors	-	10^{11} gc / mouse AAV2/8-TYMP-TYMP or AAV2/8-ALB-TYMP
Gene editing approach 2: dual rAAV2/8 vectors	Dose 1	-	4×10^{10} gc / mouse AAV2/8-Cas9 10^{11} gc / mouse AAV2/8-TYMP5 or AAV2/8-ALB2
	Dose 2	-	8×10^{10} gc / mouse AAV2/8-Cas9 2×10^{11} gc / mouse AAV2/8-TYMP5 or AAV2/8-ALB2
	Dose 3	-	$1{,}6 \times 10^{11}$ gc / mouse AAV2/8-Cas9 4×10^{11} gc / mouse AAV2/8-TYMP5 or AAV2/8-ALB2
	Dose 4	-	$3{,}2 \times 10^{11}$ gc / mouse AAV2/8-Cas9 8×10^{11} gc / mouse AAV2/8-TYMP5 or AAV2/8-ALB2

Table 19. Intravenous injection characteristics.

4.6. Experimental designs

4.6.1. *In vitro* experimentation

Several experimental designs were used, depending on the application (Fig. 27). All cell culture procedures were detailed in section 4.4. The harvested cells were processed and analysed with the methodologies detailed in sections 4.7, 4.8, 4.9 and 4.10.

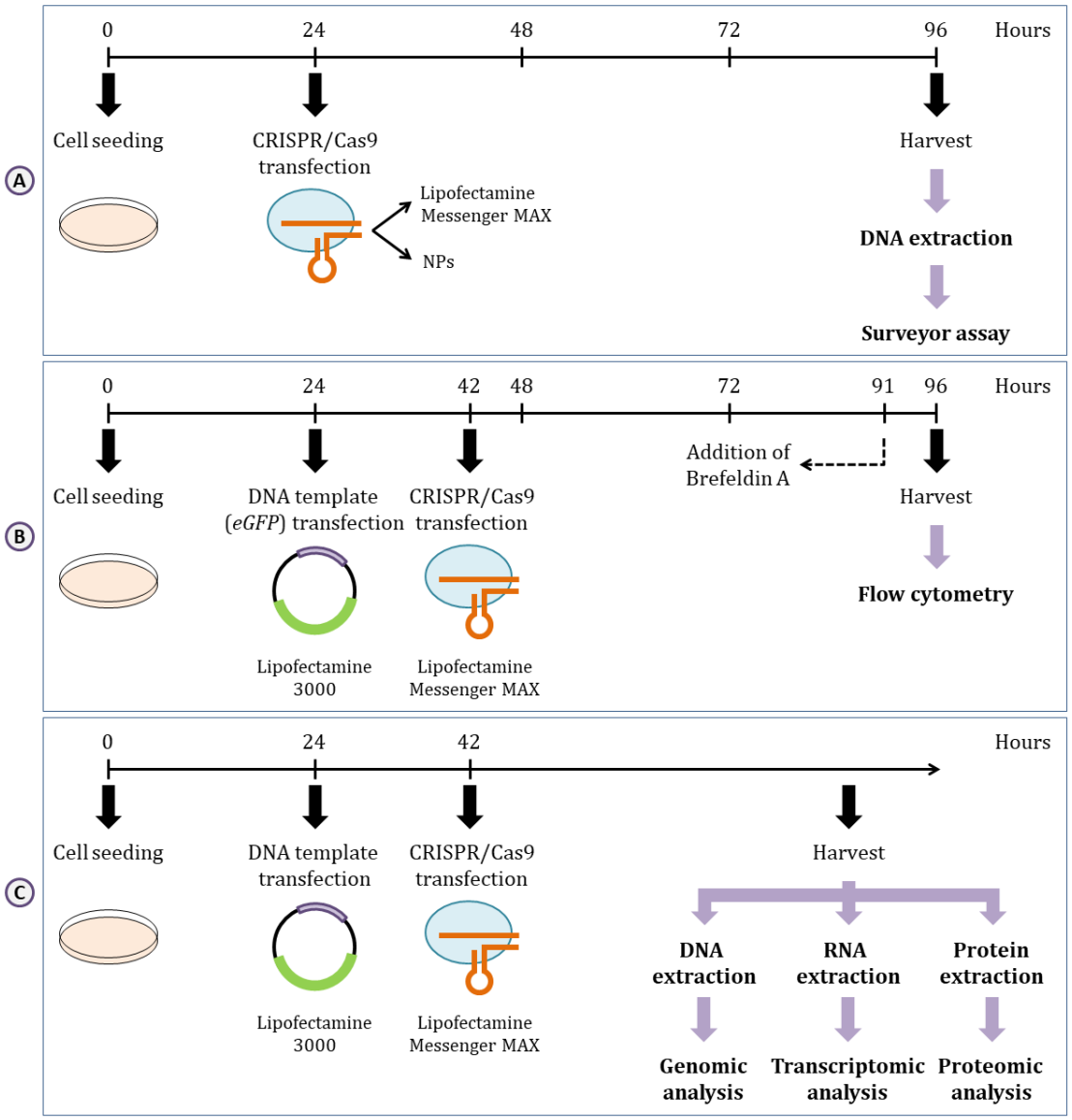

Figure 27. *In vitro* gene editing experimental designs. A) Evaluation of CRISPR/Cas9 efficiency; B) Assessment of gene editing by flow cytometry; C) Assessment of gene editing by genomic, transcriptomic and proteomic analysis.

4.6.2. *In vivo* experimentation

Several experimental designs were used, depending on the application (Fig. 28). All animal procedures were detailed in section 4.5. The blood samples and livers were processed and analysed with the methodologies detailed in sections 4.7, 4.8, 4.9 and 4.10.

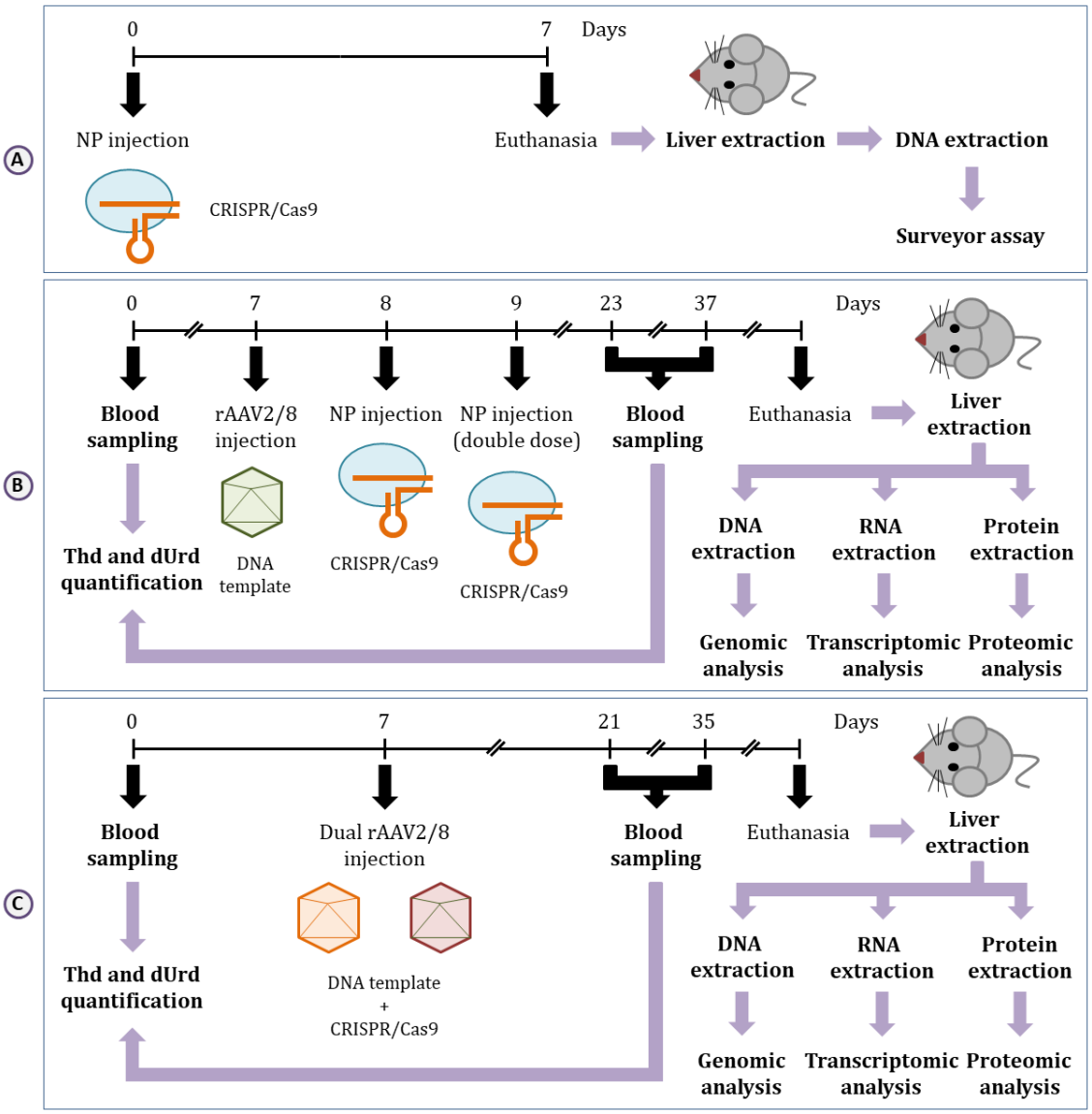

Figure 28. *In vivo* gene editing experimental designs. A) Evaluation of CRISPR/Cas9 efficiency; B) Gene editing approach 1: NPs + rAAV2/8 vectors; C) Gene editing approach 2: dual rAAV2/8 vectors.

4.7. General techniques

4.7.1. Genomic DNA extraction

Genomic DNA extraction was performed with the *GenElute Mammalian Genomic DNA Miniprep Kit* (*G1N70, Sigma-Aldrich*), following the manufacturer's instructions. Cells were lysed by different methods depending on their origin (cultured cells or liver tissue), but the genomic DNA was purified in the same manner, using silica-based membrane columns. Samples were stored at -20ºC until analysis.

4.7.2. RNA extraction

4.7.2.1. Cultured cells

RNA extraction from cultured cells was performed with the *PureLink RNA Mini Kit* (*12183025, Invitrogen*), following the manufacturer's instructions. Cells were lysed and RNA was purified using silica-based membrane columns. Samples were stored at -80ºC until analysis.

4.7.2.2. Liver

RNA extraction from tissue was performed with *RiboZol RNA Extraction Reagent* (*97064, VWR Life Science*), following the manufacturer's instructions. A liver fragment was mixed with this reagent and homogenised with the *Rotor Stator Homogenizer* (*Biokorm, Barcelona, Spain*). Then chloroform (*1.02445.1000, Merck*) was added and the sample was centrifuged at 12000 $_{x}$ g for 15 minutes at 4ºC. The upper transparent phase was separated, and then RNA was precipitated with isopropanol (*I9516, Sigma-Aldrich*). Finally, the pellet was washed with 75% ethanol (*1.08543.025, Merck*) and RNA was resuspended in nuclease-free water (*AM9906, Ambion*). Samples were stored at -80ºC until analysis.

4.7.3. Protein extraction

Protein extraction was performed by homogenising samples with a lysis buffer (0,02% 2-mercaptoethanol, 1% Triton X-100, 2 mM PMSF (phenylmethylsulfonyl fluoride), 50 mM Tris-HCl, pH 7,2). Cultured cells were homogenised by passing them through a 30 G needle several times, while liver tissues were mechanically disrupted with the *Rotor Stator Homogenizer* (*Biokorm, Barcelona, Spain*). Then, the tubes were centrifuged at 20800 $_{x}$ g for 30 minutes at 4ºC. The resulting supernatant contained the extracted proteins. Samples were stored at -20ºC until analysis.

4.7.4. DNA and RNA quantification

The *Nanodrop 2000 Spectrophotometer* (*ND-2000, Thermo Scientific*) was used to quantify and evaluate the purity of DNA and RNA. A fraction of the sample was directly loaded on the optical measurement surface, and then the nucleic acid concentration was determined.

4.7.5. Protein quantification

Protein samples were quantified with *Pierce BCA Protein Assay Kit* (*23227, Thermo Scientific*), following the manufacturer's instructions. The absorbance was measured at 562 nm in the plate reader *ELx800 Absorbance Microplate Reader* (*BioTek, Winooski, USA*). A standard curve was generated, and then the sample concentration values were interpolated using it.

4.7.6. Electrophoresis

Electrophoresis allows the separation of different nucleic acid fragments by applying an electric field to move the charged molecules through the gel. The distinct molecules are separated by their size. The reagents were a little different depending on the type of nucleic acid.

For DNA samples, the agarose gel was manufactured by dissolving *UltraPure Agarose* (*16500500, Invitrogen*) with *UltraPure TAE Buffer* (*15558026, Invitrogen*), following the manufacturer's instructions. For RNA samples, *UltraPure TBE Buffer* (*15581044, Invitrogen*) was used instead of TAE. In both cases, 0,5 µg/ml of ethidium bromide (EtBr) (*15585011, Invitrogen*), an intercalating agent, was added to be able to visualize the nucleic acids. These gels could be generated with varying percentage of agarose, which determined the precision of fragment separation. To visualize DNA (PCR products and plasmids), the standard agarose percentage used was 1%, while for RNA it was 2%. A size marker was necessary to identify each band with their corresponding size in bp. For DNA samples, the DNA marker *1 Kb Plus DNA Ladder* (*10787018, Invitrogen*) was used. For RNA samples, the size marker *RNA Century Marker* (*AM7780, Invitrogen*) was used. Both a small proportion of the nucleic acid samples and the size marker were mixed with loading dyes, following the manufacturer's instructions. The *DNA Gel Loading Dye* (*R0611, Thermo Scientific*) was used for DNA samples, while the *RNA Gel Loading Dye* (*R0641, Thermo Scientific*) was used for RNA samples. The resulting mixtures were loaded into the agarose gel, which was submerged in TAE/TBE buffer, and the electrophoresis was started. A voltage of 10 V/cm (volts per centimetre of gel length) was applied until the loading dye neared the bottom of the gel. All electrophoreses were run with the *Horizon 20-25* apparatus (*Life Technologies, Carlsbad, USA*). When the electrophoresis was finished, the gel was revealed with the UV transilluminator *ChemiDoc MP System* (*Bio-Rad, Hercules, USA*).

4.7.7. DNA purification

4.7.7.1. Gel purification

When it was necessary to isolate a particular DNA fragment or PCR product among several different fragments, gel purification was the chosen method. First, the DNA fragments were separated by size with an agarose gel electrophoresis (section 4.7.6). Then the desired band was excised and the DNA was purified with the *mi-Gel Extraction Kit* (*mi-GEL50, Metabion*), following the manufacturer's instructions. This kit provided a simple method for agarose melting and DNA purification using silica-based membrane columns. The DNA fragments were separated from the residues of the agarose gel.

4.7.7.2. Column purification

When it was not necessary to discriminate between different DNA fragments or PCR products, column purification was the chosen method. This was performed using the *mi-PCR Purification Kit* (*mi-PCR50, Metabion*), following the manufacturer's instructions. This kit provided a simple and fast DNA purification using silica-based membrane columns. The DNA fragments were separated from the enzymes, dNTPs, salts or primers used in enzymatic reactions.

4.7.8. Sequencing

4.7.8.1. Sanger sequencing

Sanger sequencing was performed at *Macrogen* (*Seoul, South Korea*) through a service provided by our institutional *Unitat d'Alta Tecnologia* (UAT). The user provided samples of 10 µl, which contained the DNA to be sequenced (250-500 ng of plasmid of PCR product) and the desired primer (2,5 µM). The resulting sequences were provided together with their respective chromatogram. Then these results were compared with the desired sequence using the software *BLASTn* (*National Center for Biotechnology Information (NCBI), Bethesda, USA*).

4.7.8.2. NGS

NGS was performed at the genomic platform of *Banc de Sang i Teixits* (BST) (*Barcelona, Spain*), thanks to the help of Dr. Francisco Vidal. The user only provided the purified PCR products. Samples were analysed in the *Illumina MiSeq* sequencing system (*Illumina, San Diego, USA*). The results were analysed with the software *QIAGEN CLC Genomics Workbench Version 20* (*QIAGEN, Hilden, Germany*). All reads were classified based on their sequence, and the frequency of each class of read was provided. This way, it was possible to quantify the presence of mutations in the desired sequence.

4.7.9. Flow cytometry

Flow cytometry was used to quantify the percentage of eGFP fluorescent cells. Depending on each experimental design, cells were harvested at different time points. The recollected cells were counted, and 2-5 $_{x}$ 10^{5} cells were deposited in each tube. Then, cells were washed with PBS-azide buffer (PBS (1x) supplemented with 0,02% of sodium azide (*S8032, Sigma-Aldrich*) and 3% bovine serum albumin (BSA) (*A7906, Sigma-Aldrich*)), and resuspended in 200-500 µl of this buffer. Finally, all samples were processed in the flow cytometer *LSRFortessa* (*BD Biosciences, San Jose, USA*). The software *FCS Express 4* (*De Novo Software, Pasadena, USA*) was used to analyse the results.

4.8. Genomic analysis

4.8.1. Assessment of CRISPR/Cas9 efficiency

4.8.1.1. Surveyor assay

The *in vitro* efficiency of CRISPR/Cas9 was assessed with the Surveyor assay. This methodology took advantage of the fact that CRISPR/Cas9 genomic DSBs were preferentially repaired by NHEJ in the absence of a template. Since it is an error-prone system, it was likely that mutations were introduced into the genomic DNA, near the gRNA target. This assay allowed the evaluation of the nuclease efficiency by quantifying the mutations caused as a side-effect.

Using the genomic DNA extracted from cultured cells as template, PCRs were performed with *Ex Taq DNA Polymerase* (*RR01AM, Takara*), following the manufacturer's instructions. The primers used allowed the amplification of fragments of the *Tymp* and *Alb* loci, which contained their respective gRNA genomic targets. The PCR conditions were the same for all reactions except the annealing temperature, which depends on the primers used; and the extension time, which depends on the size of the product (Fig. 29, Table 20). The primers were purchased from *Thermo Fisher Scientific* (*Waltham, USA*). PCR reactions were performed with *Veriti 96-Well Thermal Cycler* (*Applied Biosystems, Foster City, USA*).

Next, the pool of PCR products, which contained molecules either with or without mutations, was denaturalized and slowly reannealed using the thermocycler (Fig. 30). Some of the resulting dsDNA molecules would contain a mutated strand and a non-mutated strand, which means that they would present a mismatch.

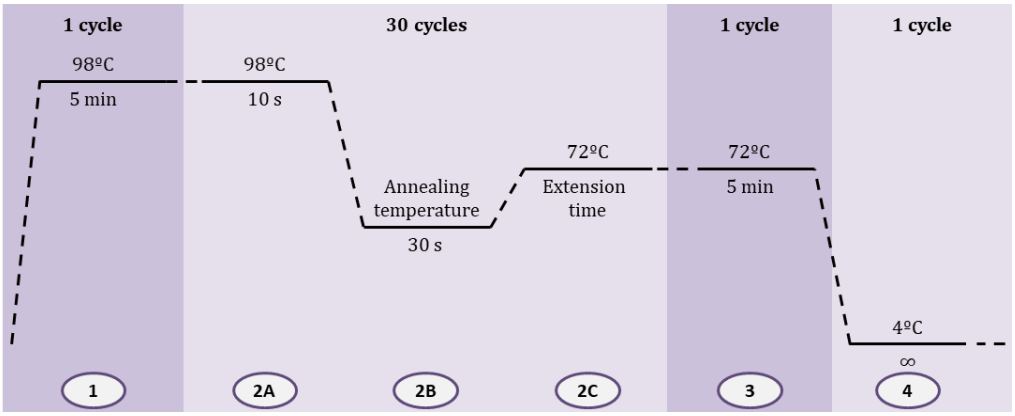

Figure 29. Surveyor assay PCR program. 1) Initial denaturalization; 2a) Denaturalization; 2b) Primer annealing; 2c) Extension; 3) Final extension; 4) Stand-by. The annealing temperatures and extension times of the different PCR reactions were specified in Table 20.

Locus	gRNA	Product size (bp)	Annealing temperature (ºC)	Extension time (s)	Primer structure (5'-3')	
Tymp	Tymp3	567	65,1	20	F	TTGACAAGTGCCGCCTACTG
					R	TGTGTCCTGTGCTCTTCCATC
	Tymp5	594	64,7	20	F	ATGAGACTGGTGAGTGCCAGAGAC
					R	CTTGTCACCCACACCTCCTGTGGA
Alb	Alb2	629	62,4	20	F	TGGGTAACCTTTCTCCTCCTCC
					R	TGGCAATGGTTCCTCTCTGCTA

Table 20. Characteristics of Surveyor assay primers and PCR program.

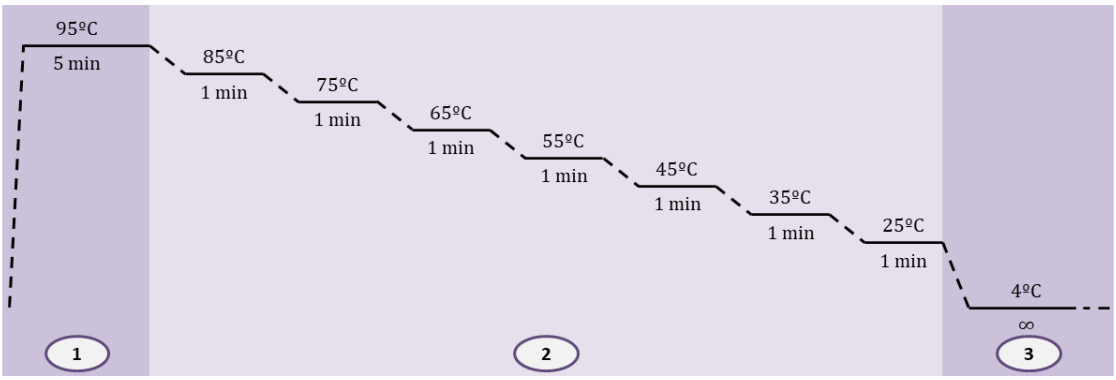

Figure 30. Surveyor assay reannealing program. 1) Initial denaturalization; 2) Reannealing: The speed of temperature decrease was: 2 ºC/s (95 → 85), 0,3 ºC/s (85 → 75), 0,3 ºC/s (75 → 65), 0,3 ºC/s (65 → 55), 0,3 ºC/s (55 → 45), 0,3 ºC/s (45 → 35), 0,3 ºC/s (35 → 25). 3) Stand-by.

Then, these dsDNA molecules were digested using the *Surveyor Mutation Detection Kit* (*706020, Integrated DNA Technologies (IDT)*), following the manufacturer's instructions. This kit contained a nuclease that has the ability to detect mismatches and digest them. Consequently, the hybrid dsDNA molecules would be cut. Finally, these processed PCR products were run in a 2% agarose gel electrophoresis (section 4.7.6). For each reaction, three bands had to be present, the parental band and the digested bands (Fig. 31). The intensity of all bands was quantified using the software *ImageJ* (*National Institute of Health (NIH), Bethesda, USA*). The relative amount of digested bands was used to determine the nuclease efficiency, using the following formula:

$$\text{CRISPR/Cas9 Efficiency (\%)} = \left[1 - \sqrt{1 - \frac{\sum \text{Intensity of digested bands}}{\sum \text{Intensity of all bands}}}\right] \times 100$$

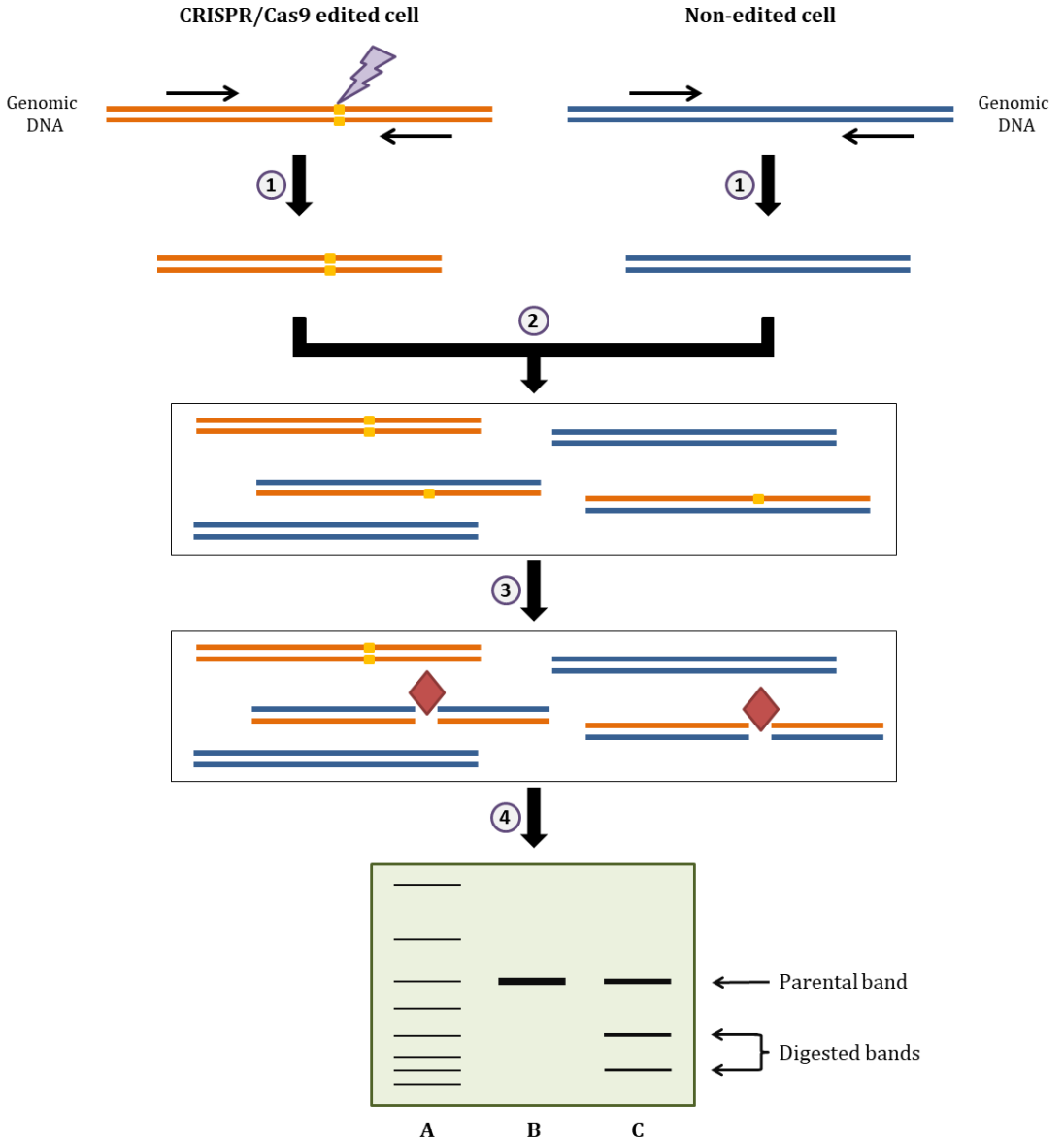

Figure 31. Schematic representation of the Surveyor assay. 1) PCR with specific primers that allowed the amplification of the CRISPR/Cas9 genomic targets. 2) Denaturalization and reannealing of the PCR products. 3) Digestion of hybrid dsDNA molecules with mismatches. 4) Electrophoresis of the digested PCR products. Lane A: size marker; lane B: C- (non-transfected cells); lane C: sample (cells transfected with CRISPR/Cas9). For Tymp3 gRNA, the size of the parental band was 567 bp, and the size of the digested bands was 389 bp and 178 bp. For Tymp5 gRNA, the size of the parental band was 594 bp, and the size of the digested bands was 345 bp and 249 bp. For Alb2 gRNA, the size of the parental band was 629 bp, and the size of the digested bands was 395 bp and 234 bp. The primer sequences were specified in Table 20. C-: negative control.

4.8.1.2. NGS

The *in vivo* efficiency of CRISPR/Cas9 was assessed by NGS. Using the genomic DNA extracted from livers as template, PCRs were performed with *Ex Taq DNA Polymerase* (*RR01AM, Takara*), using the same primers and conditions mentioned in the previous section. Then the PCR products were purified (section 4.7.7.2) and sent to BST to be sequenced (section 4.7.8.2). The nuclease efficiency was calculated through the proportion of reads with mutations near the gRNA target sequence, using the following formula:

$$\text{CRISPR/Cas9 Efficiency (\%)} = \sum \text{Frequency of reads with mutations}$$

4.8.2. Detection of gene editing

It was possible to ascertain the presence of edited cells through the detection of the inserted DNA templates. Using the genomic DNA extracted from either cultured cells or livers, PCR reactions with specific primers were performed (Table 21). All PCRs were done with *Phusion DNA Polymerase* (*F530, Thermo Scientific*), following the manufacturer's instructions. The PCR conditions were the same for all reactions except the annealing temperature, which depends on the primers used; and the extension time, which depends on the size of the product (Fig. 23, Table 22). All primers were purchased from *Thermo Fisher Scientific* (*Waltham, USA*). The apparatus used to perform all reactions was *Veriti 96-Well Thermal Cycler* (*Applied Biosystems, Foster City, USA*). The PCR products were run in a 1% agarose gel electrophoresis (section 4.7.6). Through the size of the observed bands, it was possible to discriminate between edited and non-edited DNA molecules. Moreover, the size of the bands allowed the differentiation of the insertion mechanism (HDR or NHEJ). The intensity of all bands was quantified using the software *ImageJ* (*NIH, Bethesda, USA*). The relative intensity of edited bands was used to estimate the gene editing efficiency, using the following formula:

$$\text{Gene Editing Efficiency (\%)} = \left[1 - \sqrt{1 - \frac{\sum \text{Intensity of edited bands}}{\sum \text{Intensity of all bands}}}\right] \times 100$$

In some cases the gene editing was undetectable with a simple PCR due to its low efficiency. To solve this, nested PCRs with specific primers were performed (Fig. 23, Table 21, 22). The first PCR products were diluted 1:50 in water, and they were used as templates for a second PCR. With this methodology, it was possible to detect edited bands in the agarose gel. In these cases, the gene editing efficiency could not be quantified with the previous formula, but it was possible to ascertain the presence of edited cells.

The PCR strategy was different for *eGFP* cDNA (Fig. 32, Table 21, 22), and m*Tymp*/h*TYMP* cDNAs (Fig. 33, Table 21, 22). The simple PCR for m*Tymp*/h*TYMP* cDNAs allowed the differentiation between edited and unedited DNA molecules, as the primers were located outside the insertion region. But the strategy for *eGFP* cDNA only allowed the differentiation between the two insertion methods, and unedited molecules were not

detectable. For this reason, the gene editing efficiency was quantifiable only for m*Tymp*/h*TYMP* cDNAs. The insertion of *eGFP* cDNA was detectable but not quantifiable.

In all cases, the bands corresponding to HDR and NHEJ insertion were extracted from the agarose gel and purified (section 4.7.7.1). Their sequence was verified by Sanger sequencing (section 4.7.8.1).

Locus	cDNA	Type of PCR	Primer structure (5'-3')	
Tymp	*eGFP*	Simple	F	CCTAGAGATCTGTCGCTTGCC
			R	AAGCTGCAATAAACAAGTAGGGATCCTTACTTGTACAG
		Nested	F	TCCGGAGCGCTGTAACCCGA
			R	CGAACTTCACCTCGGCGCG
	m*Tymp*	Simple	F	ATGAGACTGGTGAGTGCCAGAGAC
			R	AGCGAGGTGTTCACGCAGTC
		Nested	F	ATGAGACTGGTGAGTGCCAGAGAC
			R	TCAACAATGCCGTCTGCCGGTGC
Alb	*eGFP*	Simple	F	TGGGTAACCTTTCTCCTCCTCC
			R	AAGCTGCAATAAACAAGTAGGGATCCTTACTTGTACAG
		Nested	F	CGCCGAGAAGCACGTAAGAG
			R	TGAAGAAGTCGTGCTGCTTC
	h*TYMP*	Simple	F	TGGGTAACCTTTCTCCTCCTCC
			R	AGATCCACAGTCTTTCTGTC
		Nested	F	CGCCGAGAAGCACGTAAGAG
			R	GACCAGCTCCACGGTGCCATC

Table 21. Characteristics of the primers used to detect gene editing.

Locus	cDNA	Type of PCR	Annealing temperature (ºC)	Extension time (s)	Expected product size (bp)		
					Unedited	HDR insertion	NHEJ insertion
Tymp	*eGFP*	Single	65,1	45	-	-	~1600
		Nested	68,6	30	-	-	~1100
	m*Tymp*	Single	67,8	45	~400	-	~1800
		Nested	68,7	30	-	-	~1300
Alb	*eGFP*	Single	66,1	40	-	~1200	~1500
		Nested	63,5	20	-	~600	~900
	h*TYMP*	Single	59,8	60	~1000	~2600	~3300
		Nested	65,7	45	-	~1500	~1800

Table 22. PCR program characteristics for detecting gene editing.

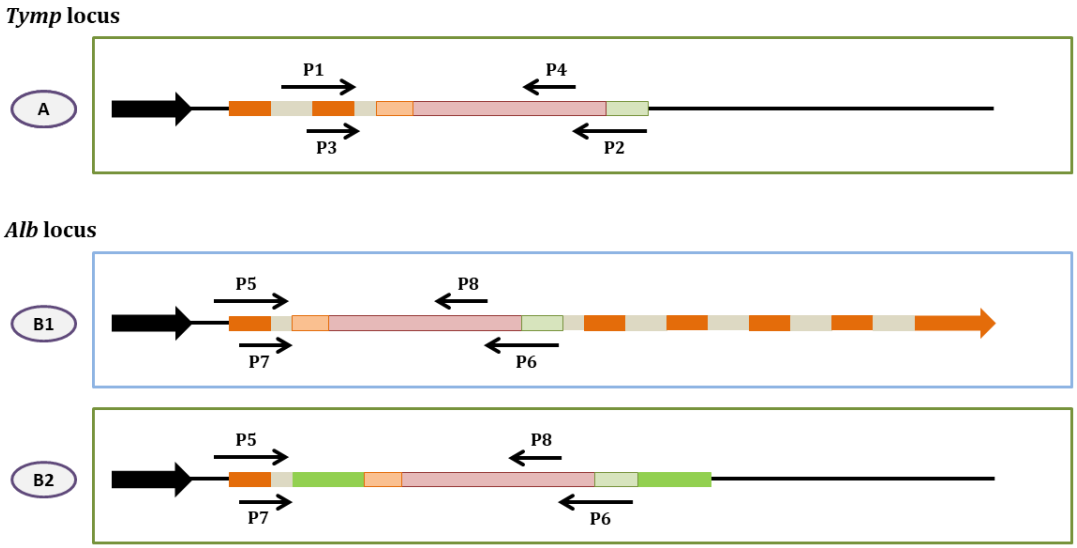

Figure 32. Schematic representation of the strategy to detect the inserted *eGFP* cDNA. *Tymp* locus: a PCR with primers P1 and P2 allowed the detection of molecules edited by NHEJ (A). A nested PCR with primers P3 and P4 allowed the further amplification these edited molecules. *Alb* locus: the size of the PCR products generated with primers P5 and P6 allowed the discrimination between molecules edited by HDR (B1) or NHEJ (B2). A nested PCR with primers P7 and P8 allowed the further amplification of these edited molecules. The primer sequences were specified in Table 21. The size of all bands was specified in Table 22.

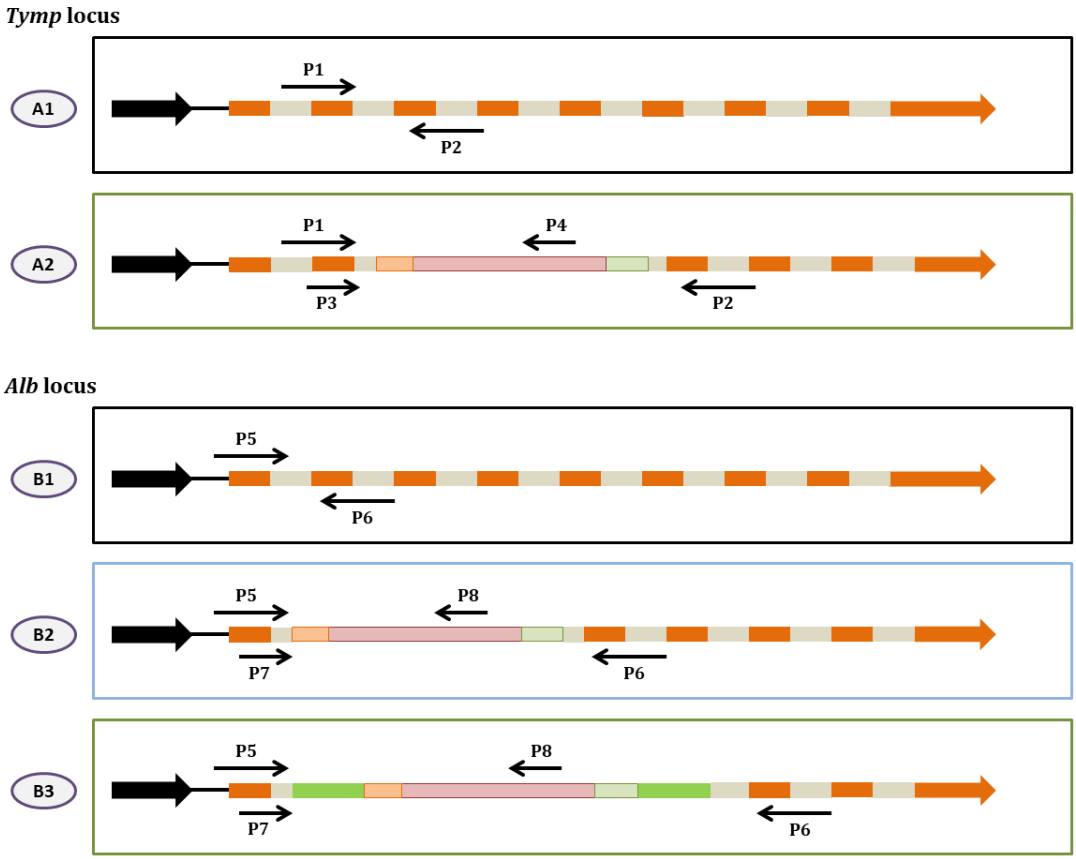

Figure 33. Schematic representation of the strategy to detect the inserted m*Tymp*/h*TYMP* cDNAs. *Tymp* locus: the size of the PCR products generated with primers P1 and P2 allowed the discrimination between molecules unedited (A1) and edited by NHEJ (A2). A nested PCR with primers P3 and P4 allowed the further amplification of molecules edited by NHEJ (A2). *Alb* locus: the size of the PCR products generated with primers P5 and P6 allowed the discrimination between molecules unedited (B1), and edited by HDR (B2) or NHEJ (B3). A nested PCR with primers P7 and P8 allowed the further amplification of molecules edited by HDR (B2) or NHEJ (B3). The primer sequences were specified in Table 21. The size of all bands was specified in Table 22.

4.9. Transcriptomic analysis

4.9.1. Quantification of TP mRNA

The first step was the reverse transcription of RNA samples using the *High-Capacity cDNA Reverse Transcription Kit* (*4368814, Applied Biosystems*), following the manufacturer's instructions. The apparatus used was *Veriti 96-Well Thermal Cycler* (*Applied Biosystems, Foster City, USA*) (Fig. 34).

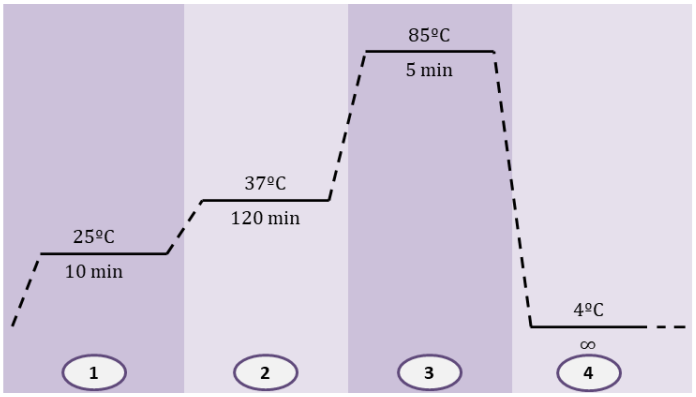

Figure 34. Reverse transcription program. 1) Initial step; 2) Reverse transcription; 3) Enzymatic inactivation; 4) Stand-by.

The amount of TP mRNA (converted to cDNA) was quantified by real-time PCR using the *TaqMan Fast Advanced Master Mix* (*4444556, Applied Biosystems*), following the manufacturer's instructions. This methodology required the use of custom-made *TaqMan Gene Expression Assays* (*4448892, Applied Biosystems*), which consisted of a pair of unlabeled specific primers and a TaqMan probe with a dye label (Fluorescein amidite, FAM) on the 5'-end and a minor groove binder (MGB) and non-fluorescent quencher (NFQ) on the 3'-end. Two *TaqMan Gene Expression Assays* were used, and each one allowed the quantification of a specific cDNA. One was used to quantify h*TYMP* cDNA (Hs00157317_m1), and the other was used to quantify m*Gapdh* cDNA (Mm99999915_g1), which served as an internal housekeeping expression control (calibrator). For each cDNA sample, two different tests had to be performed, one with h*TYMP* cDNA and another with the calibrator. Each test was run in duplicates. All reactions were performed with the *ABI Prism 7000 Thermal Cycler* (*Applied Biosystems, Foster City, USA*) (Fig. 35).

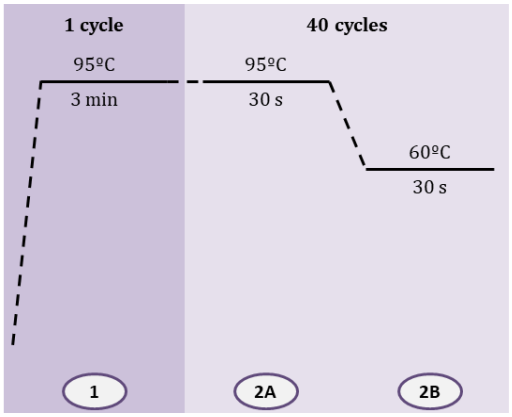

Figure 35. Real-time PCR program. 1) Initial denaturalization; 2a) Denaturalization; 2b) Primer annealing and extension.

For each assay, the primers allowed the amplification of a fragment of the cDNA sequence, while the TaqMan probe annealed to a sequence located between both primers. In each cycle of the PCR program, the DNA polymerase would hydrolyse the probe during its action. As a result, the FAM label would be separated from the NFQ-MGB tags, which were blocking the fluorescence. In each cycle, the number of hydrolysed probes would increase, and at a certain point the fluorescence would become detectable. The cycle number where fluorescence started being detected was called threshold cycle (Ct). This parameter was used to quantify the relative amount of a cDNA, as a lower Ct implied a higher presence of cDNA in the starting sample and vice versa. The results of the real-time quantitative PCR were shown as a graph that correlated the fluorescence intensity with the cycle number of the program (Fig. 36). The Ct parameter was determined in all tests, and then the hTP mRNA expression was calculated as a normalized ratio. This parameter was obtained using the following formula:

$$\text{hTP mRNA expression ratio} = 2^{[Ct\ (\text{calibrator}) - Ct\ (\text{sample})]}$$

The mTP mRNA could not be quantified with this technique as there was not a suitable *TaqMan Gene Expression Assay* available. For this reason, only the mRNA of *Alb* locus gene editing was analysed and quantified.

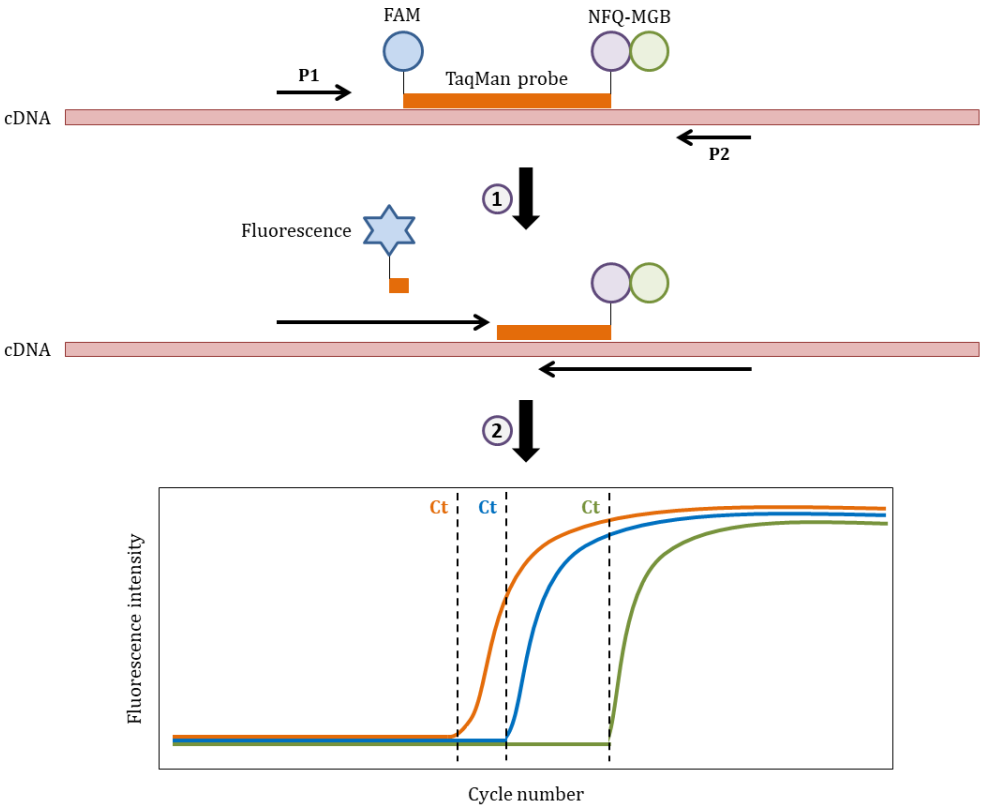

Figure 36. Schematic representation of real-time quantitative PCR. 1) The amplification of a cDNA fragment through a real-time PCR with primers P1 and P2 provoked the hydrolysis of the TaqMan probe. This caused the liberation of FAM, so fluorescence was emitted. 2) In each cycle of the PCR program the fluorescent molecules accumulated until they reached a detectable threshold. The cycle where this happened was named Ct.

4.10. Proteomic analysis

4.10.1. Detection of cells edited with *eGFP* cDNAs

The efficiency of *in vitro* gene editing experiments using the *eGFP* cDNA was assessed by flow cytometry (section 4.7.9). Transfected cells were recollected and analysed at the cytometer. The percentage of fluorescent cells was used to quantify the gene editing efficiency, as the number of fluorescent cells correlated with the number of edited cells.

4.10.2. Detection of TP protein

TP protein was detected and quantified by Western blot. The first step was the polyacrylamide gel electrophoresis, which allowed the separation of different proteins by their weight. The pre-made *NuPAGE 4-12% Bis-Tris Gels* (*NP0321, Invitrogen*) were used. Gels were submerged in *NuPAGE MOPS SDS Running Buffer* (*NP0001, Invitrogen*), which was supplemented with *NuPAGE Antioxidant* (*NP0005, Invitrogen*) in the internal chamber. Protein samples were mixed with *NuPAGE LDS Sample Buffer* (*NP0007, Invitrogen*) and *NuPAGE Sample Reducing Agent* (*NP0004, Invitrogen*), following the manufacturer's instructions. These samples were incubated at 70ºC for 10 minutes to be denaturalized. The size marker *PageRuler Plus Prestained Protein Ladder* (*26619, Thermo Scientific*) was loaded on the first well of the gel. The other wells were loaded with 10-50 µg of total protein samples. A voltage of 200 V was applied until the loading dye neared the bottom of the gel. All electrophoresis were run with the apparatus *XCell SureLock Mini-Cell* (*Invitrogen, Carlsbad, USA*).

When the electrophoresis was finished, the gel was extracted, placed inside *iBlot 2 PVDF Mini Stacks* (*IB24002, Invitrogen*), and loaded into the *iBlot 2 Gel Transfer Device* (*Invitrogen, Carlsbad, USA*). A voltage of 25 V and an amperage of 2,5 A were applied for 7 minutes. This way the proteins were transferred into a PVDF membrane.

The next step was the immunoblotting. The PVDF membrane was washed by being submerged in TBS-T buffer (1x) and incubated at RT in a shaking device. Every 5-10 minutes the TBS-T (1x) was renewed. A minimum of three washes were performed this way. Next, the membrane was blocked by being submerged in blocking buffer (1x) (TBS-T (1x) supplemented with 5% dry milk (*9999S, Cell Signalling Technology (CST)*). This incubation was done at RT for 1 hour in a shaking device. Then, the membrane was washed at least three times with TBS-T (1x), like mentioned above. Next, the membrane was incubated with the primary antibody, which was diluted in blocking buffer (1x) (Table 23). This incubation was done at 4ºC overnight in a shaking device. The next day the membrane was washed at least three times with TBS-T (1x). Then, it was incubated with a secondary molecule, which was diluted in blocking buffer (1x) (Table 23). This incubation was done at RT for 1 hour in a shaking device. Finally, the membrane was washed at least three times with TBS-T (1x).

Species	Primary incubation		Secondary incubation	
	Molecule	Dilution	Molecule	Dilution
Mouse TP	Polyclonal antibody Anti-TP IgG from Sheep (*AF7568, Research and Diagnostic Systems (R&D)*)	1/2000	Secondary antibody Anti-Sheep IgG conjugated with HRP (*A16041, Invitrogen*)	1/10000
Human TP	Monoclonal antibody Anti-TP IgG_1 Clone P-GF.44C from Mouse conjugated to Biotin (*NB100-2737B, Novus Biologicals*)	1/500	Streptavidin conjugated with HRP (*21130, Thermo Scientific*)	1/5000

Table 23. Immunoblotting characteristics.

The last step was revealing the membrane by incubating it with *Pierce ECL Western Blotting Substrate* (*32109, Thermo Scientific*) at RT for 5 minutes in the dark. This reagent, which contained horseradish peroxidase (HRP) substrate, reacted with the HRP molecules attached to TP. The enzymatic reaction of HRP produced chemiluminescence, which was detected with *Odyssey Fc Imaging System* (*LI-COR Biosciences, Lincoln, USA*). The detection of a band of approximately 55 kDa revealed the presence of TP in samples.

To be able to relatively quantify TP it was necessary to use a housekeeping protein as a loading control. In this case β-actin was used, as its expression in the liver was stable. To this end, the membrane was washed by being submerged in stripping buffer (1x) and incubated at RT in a shaking device. Every 5-10 minutes the stripping buffer (1x) was renewed. A minimum of three washes were performed this way. This buffer allowed the elimination of antibodies attached to the PDVF membrane. Next, the membrane was washed two times with PBS (1x), and then two times with TBS-T (1x). At this point, the membrane was ready for another immunoblotting. The membrane was blocked and antibody-stained as detailed above, but this time the primary incubation was done with the monoclonal antibody *Anti-β-actin IgG_2A Clone AC-74 from Mous*e (*A5316, Sigma-Aldrich*), and the secondary incubation was done with the polyclonal antibody *Anti-mouse IgG from Goat conjugated with HRP* (*A90-105P, Bethyl Laboratories*). The membrane was also revealed as mentioned above. This time, a band of 42 kDa had to be detected in all samples.

Finally, TP was quantified with a normalized expression ratio. First, the intensity of all bands was quantified using the software *ImageJ* (*NIH, Bethesda, USA*). For each sample, the intensity of TP band was normalized with the intensity of its respective β-actin band, using the following formula:

$$\text{Normalized ratio} = \frac{\text{TP band intensity}}{\beta\text{–actin band intensity}}$$

The final TP expression values were calculated by using the following formula:

$$\text{TP expression (\%)} = \frac{\text{Sample normalized ratio}}{\text{Control normalized ratio}} \times 100$$

The mTP protein was only quantified in the *in vivo* experiments. Since the Hepa1c1c7 cell line had a basal mTP expression, the *in vitro Tymp* locus gene editing approach did not allow the differentiation between edited and unedited cells by their mTP protein amounts. Since dKO mice did not have basal mTP expression, it was possible to detect successful *in vivo Tymp* locus gene editing. The hTP protein was quantified in both *in vitro* and *in viv*o experiments.

For the *in vitro* experiments, the band intensity ratio of cells transfected with the p305-TP plasmid was used as control. Consequently, non-treated cells had a 0% TP expression, while cells transfected with this control plasmid had a 100% TP expression. For the *in vivo* experiments, the band intensity ratio of WT mice was used as control. Consequently, non-treated dKO mice had a 0% TP expression, while WT mice had a 100% TP expression.

TBS buffer (10x) was manufactured by mixing 87,66 g of NaCl (*S3014, Sigma-Aldrich*) and 24,23 g of Tris (*154563, Sigma-Aldrich*) in 1 L of water. The pH was adjusted to 7,6. The final TBS-T buffer (1x) was obtained by mixing 900 ml of water, 100 ml of TBS (10x), and 1 ml of Tween-20 (*P7949, Sigma-Aldrich*). The stripping buffer (1x) was manufactured by mixing 15 g of glycine (*G8898, Sigma-Aldrich*), 1 g of sodium dodecyl sulfate (SDS) (*L4509, Sigma-Aldrich*) and 10 ml of Tween-20 (*P7949, Sigma-Aldrich*) in 0,99 L of water. The pH was adjusted to 2,2.

4.10.3. Determination of TP activity

Protein samples from cultured cells were diluted to obtain a concentration of 0,135 µg/µl, while the ones from liver tissues were diluted to obtain a concentration of 1,35 µg/µl. In all assays, a positive control with known TP activity was added. A lymphoblastoid cell lysate (5×10^6 cells) was used as control. For each sample, two different reactions were prepared, one for determination, and another as blank (Table 24).

	Determination	**Blank**
Sample	74 µl	74 µl
Tris-arsenate buffer (0,1 M, pH 6,5)	20 µl	20 µl
Thd (166 mM) (*T9250, Sigma-Aldrich*)	6 µl	-
Total volume	100 µl	94 µl

Table 24. Reactions for TP activity determination.

The reactions were incubated at 37ºC for 1 hour. In this step, Thd was converted to Thy through the action of TP present in the samples. Consequently, a higher TP activity correlated with a higher Thy generation. Then, 1 ml of $HClO_4$ (0,55 M) was added and the tubes were incubated in ice for 5 minutes. With this, TP enzymatic reactions were stopped and the proteins were precipitated. Then, 6 µl of Thd (166 mM) was added to blank reactions, and the tubes were further incubated on ice for 5 minutes. Finally, the tubes were centrifuged at 20800 $_x$ g for 10 minutes at 4ºC. The resulting supernatants, which contained the Thy molecules, were analysed by high performance liquid chromatography (HPLC) with ultraviolet spectrophotometric detection. The absorbance of the eluted samples was monitored at 267 nm. In each run, multi-standards with known Thy concentrations were also analysed. At the end, the area of the peak corresponding to Thy was determined in all samples. A standard curve was generated with the multi-standard values, and then the sample concentration values were interpolated using it. For each sample, the blank reaction value was subtracted from the determination reaction value. Finally, TP activity was calculated using the following formula:

$$\text{TP activity}\left(\frac{\text{nmol Thy}}{\text{h} \times \text{mg protein}}\right) = \frac{\text{Thy concentration (µM)} \times 1{,}1\ \text{(ml)}}{1\ \text{(h)} \times 0{,}072\ \text{(ml)} \times \text{Initial protein concentration (mg/ml)}}$$

This procedure was performed with the apparatus *Acquity UPLC* (Waters Corporation, Milford, USA). Ferran Vila, a PhD student employed in Dr. R. Martí research group, kindly manipulated this apparatus and analysed all samples.

4.10.4. Quantification of plasma Thd and dUrd

For each diluted plasma sample (1:9), two different reactions were prepared, one for determination, and another as blank (Table 25).

	Determination	**Blank**
Sample	200 µl	200 µl
Purified TP (1397 U/ml) (*T2807, Sigma-Aldrich*)	-	2 µl
Total volume	202 µl	200 µl

Table 25. Reactions for Thd and dUrd quantification.

The reactions were incubated at 37ºC for 30 minutes. In the blank reactions, Thd and dUrd were catabolised through the action of TP. Then, the proteins were precipitated by adding 9 µl of $HClO_4$ (70%) in all tubes and incubating them in ice for 10 minutes. Finally, the tubes were centrifuged at 20800 $_x$ g for 10 minutes at 4ºC. The resulting supernatants, which contained the Thd and dUrd molecules, were analysed by HPLC with ultraviolet spectrophotometric detection. This analysis was the same as the one mentioned in the

previous section, but this time the areas of the peaks that corresponded to Thd and dUrd were determined. The multi-standards were also used to generate a standard curve for each deoxyribonucleotide, and then the sample concentration values were interpolated using them. For each sample, the blank reaction value was subtracted from the determination reaction value, and the resulting value was multiplied by the dilution factor.

4.11. Statistical analysis

All statistical analyses were performed with the software *GraphPad Prism 6* (*GraphPad Software Inc., California, USA*). Non-parametric tests were used, as the sample size was small and a normal distribution could not be assumed. Kruskal-Wallis test was used with unpaired values, while Friedman test was used with paired measures. In all cases the Dunn's multiple comparisons post hoc test was performed. The particular test used in each case was specified in each figure. The 0,05 p-value was used as a threshold for statistical significance.

5. RESULTS

5.1. *In vitro* experimentation

5.1.1. CRISPR/Cas9 efficiency

First of all, we screened the *Tymp* and *Alb* loci searching for PAM motifs using the bioinformatics tool developed by *Zhang Lab* (*MIT, Cambridge, USA*). This software delivered a list of potential CRISPR/Cas9 targets within those loci. Each target of the list had a score associated that reflected their probability of producing off-target cleavage. For each locus, we choose the 3-5 targets with the highest score (lower off-target cleavage risk) and generated their corresponding gRNA molecules. Then, we transfected AML-12 cells with each different gRNA (and Cas9 mRNA) and performed Surveyor assays (data not shown). We selected Alb2 gRNA for *Alb* locus, based on its higher level of cutting efficiency. Similarly, Tymp3 gRNA was first chosen for *Tymp* locus, but since the gene editing approach was modified later, we selected Tymp5 gRNA as a substitute (Fig. 37a).

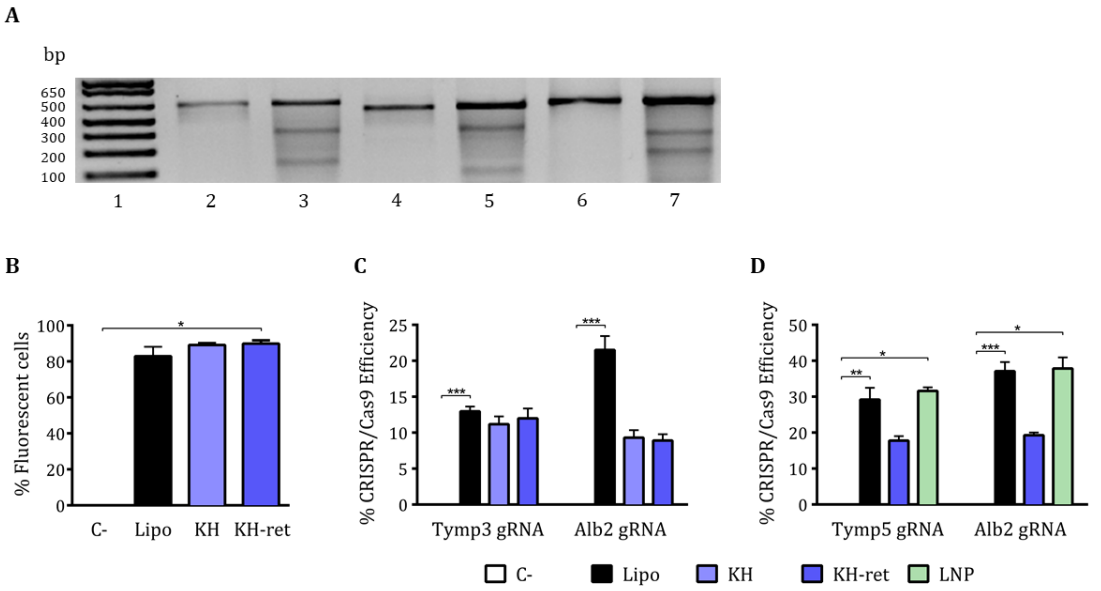

Figure 37. *In vitro* CRISPR/Cas9 cutting efficiency results. A) Representative image of Surveyor assay results. Lane 1: Size marker; lane 2: C- of Alb2 gRNA (non-transfected cells); lane 3: Alb2 gRNA sample (transfected cells); lane 4: C- of Tymp3 gRNA (non-transfected cells); lane 5: Tymp3 gRNA sample (transfected cells); lane 6: C- of Tymp5 gRNA (non-transfected cells); lane 7: Tymp5 gRNA sample (transfected cells). **B)** Transfection efficiency of eGFP mRNA in AML-12 cells. **C)** CRISPR/Cas9 efficiency in AML-12 cells. **D)** CRISPR/Cas9 efficiency in Hepa1c1c7 cells. Error bars represent the standard error of the mean (SEM). Data analysed with Kruskal-Wallis test with Dunn's multiple comparisons post hoc test. * = p<0,05; ** = p<0,01; *** = p<0,001.

Initially, we evaluated several PNP formulations carrying the eGFP mRNA in AML-12 cells (data not shown). Based on their higher transfection efficiency, we choose two PNP formulations, KH and KH-ret. Their transfection efficiency was equal to that obtained with *Lipofectamine Messenger MAX* (Fig. 37b, Table 26). Next, we tested both formulations carrying the CRISPR/Cas9 system RNAs in AML-12 cells. In this case, different results were obtained depending on the locus. While the CRISPR/Cas9 cutting efficiency (measured by

the Surveyor assay) in the *Tymp* locus was similar between *Lipofectamine Messenger MAX*, KH, and KH-ret, in the *Alb* locus the efficiency with *Lipofectamine Messenger MAX* was superior to KH and KH-ret, although these differences were not statistically significant (Fig. 37c, Table 26). Since the KH-ret formulation was the one chosen for *in vivo* experimentation, we also evaluated this PNP in Hepa1c1c7 cells. In this case, the CRISPR/Cas9 efficiency with *Lipofectamine Messenger MAX* was superior in both loci, although these differences were not statistically significant (Fig. 37d, Table 26).

We also tested LNPs carrying the CRISPR/Cas9 system *in vitro* in Hepa1c1c7 cells. In both loci, their cutting efficiency was equivalent to that obtained with *Lipofectamine Messenger MAX*, and significantly different from the negative control (Fig. 37d, Table 26).

		C-	*Lipofectamine Messenger MAX*	KH	KH-ret	LNP
Fluorescence (%) [AML-12]		0 (n=3)	82,9 ± 3,01 (n=3)	89,14 ± 056 (n=3)	89,9 ± 1,03 (n=3)	-
CRISPR/Cas9 efficiency (%) [AML-12]	**Tymp3**	0 (n=8)	12,96 ± 0,67 (n=8)	11,17 ± 1,08 (n=3)	11,98 ± 1,4 (n=3)	-
	Alb2	0 (n=22)	21,52 ± 1,92 (n=22)	9,3 ± 1,03 (n=5)	8,91 ± 0,87 (n=3)	-
CRISPR/Cas9 efficiency (%) [Hepa1c1c7]	**Tymp5**	0 (n=8)	29,19 ± 3,28 (n=8)	-	17,78 ± 1,24 (n=3)	31,6 ± 1 (n=3)
	Alb2	0 (n=10)	37,09 ± 2,54 (n=10)	-	19,26 ± 0,77 (n=3)	37,86 ± 3,09 (n=3)

Table 26. *In vitro* CRISPR/Cas9 cutting efficiency results. Values represent the mean percentage (%) ± SEM. n: sample size.

5.1.2. rAAV2/8 expression kinetics

Prior to establishing a rAAV2/8 expression kinetics *in vitro*, a range of different escalating MOIs were used initially to establish the optimal one (data not shown). It is a known that AAV vectors have poor *in vitro* transduction efficiency in comparison to that observed *in vivo*. For this reason, only the highest MOI (10^5) resulted in measurable eGFP expression. Using this MOI, we monitored the percentage of fluorescent cells for one week after transduction: 1,09 ± 0,17 % at day 3; 3,55 ± 0,75 % at day 4; 6,77 ± 0,71 % at day 5; 18,3 ± 1,51 % at day 6; and 12,34 ± 2,63 % at day 7. Consequently, we stablished that the peak of expression takes place approximately at day 6 post-transduction (Fig. 38).

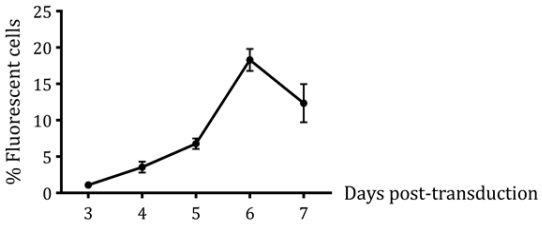

Figure 38. *In vitro* expression kinetics of AAV2/8-eGFP vector in AML-12 cells. Error bars represent the SEM (n=3 in all time points).

5.1.3. Gene editing assessment

5.1.3.1. *eGFP* gene editing

5.1.3.1.1. Genomic insertion

Through PCRs with specific primers it was possible to ascertain the insertion of *eGFP* DNA templates in the correct position. The sequence of the bands was verified by Sanger sequencing.

For *Tymp* locus, the band corresponding to NHEJ insertion was not visible in a simple PCR (Fig. 39a), but it was detectable with a nested PCR (Fig. 39b). This means that the gene editing efficiency was low. For *Alb* locus, the bands corresponding to HDR and NHEJ were visible both with simple and nested PCRs (Fig. 39). In both cases the HDR band was much more intense, which means that the gene editing by HDR was more efficient *in vitro* in Hepa1c1c7 cells.

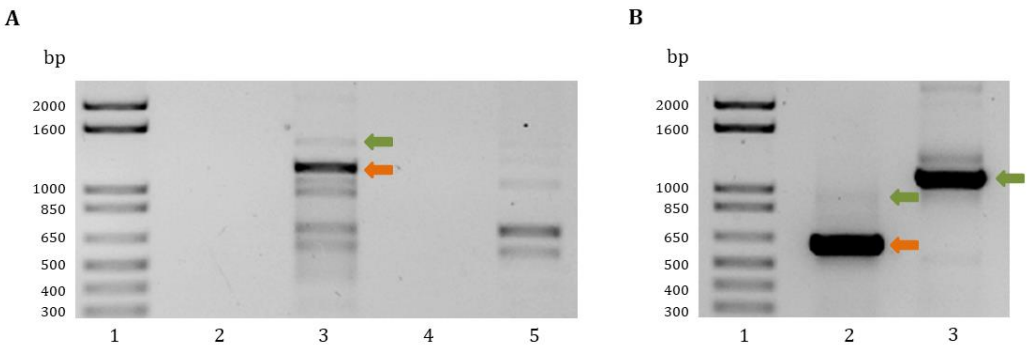

Figure 39. Detection of genomic insertion of the *eGFP* DNA templates. A) Representative image of simple PCR to detect gene editing. Lane 1: Size marker; lane 2: C- of *Alb* locus; lane 3: *Alb* locus gene editing sample; lane 4: C- of *Tymp* locus; lane 5: *Tymp* locus gene editing sample. **B)** Representative image of nested PCR to detect gene editing. Lane 1: Size marker; lane 2: *Alb* locus gene editing sample (transfected cells); lane 3: *Tymp* locus gene editing sample (transfected cells). Green arrows point to NHEJ bands, while orange arrows point to HDR bands.

5.1.3.1.2. *eGFP* expression

The *eGFP* cDNA was used to easily detect edited Hepa1c1c7 cells. Cells transfected with only the DNA templates were used as intrinsic controls to discard the background fluorescence produced by the template plasmids alone. They were also transfected with pEGFP-1 plasmid and eGFP mRNA to ascertain the transduction efficiency of both DNA templates and CRISPR/Cas9 system.

At both loci the percentage of fluorescence of cells transfected with both DNA templates and CRISPR/Cas9 system were higher than in those transfected with the templates alone, suggesting that a small proportion of cells were correctly edited. At the *Tymp* locus this difference was not significant (Fig. 40a, Table 27), but at the *Alb* locus the difference was highly significant (Fig. 40b, Table 27). In both cases, the percentage of fluorescence was significantly different from the negative control (non-transfected cells) (Fig. 40).

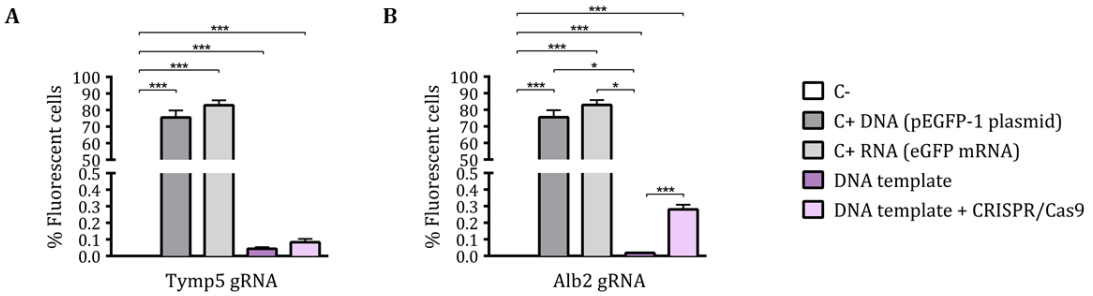

Figure 40. *eGFP* expression in edited cells. Percentage of fluorescence in cells edited with Tymp5 gRNA (**A**) and Alb2 gRNA (**B**). Error bars represent the SEM. Data analysed with Kruskal-Wallis test with Dunn's multiple comparisons post hoc test. * = p<0,05; *** = p<0,001. C+: positive control.

		C-	DNA template	DNA template + CRISPR/Cas9	C+ DNA (pEGFP-1)	C+ RNA (eGFP mRNA)
Fluorescence (%)	Tymp5	0 (n=18)	0,044 ± 0,01 (n=18)	0,083 ± 0,019 (n=18)	75,5 ± 4,31 (n=3)	82,9 ± 3,01 (n=3)
	Alb2	0 (n=40)	0,019 ± 0,001 (n=40)	0,281 ± 0,028 (n=40)		

Table 27. *eGFP* expression in edited cells. Values represent the mean % ± SEM.

5.1.3.2. TP gene editing

5.1.3.2.1. Genomic insertion

Through PCRs with specific primers it was possible to ascertain the insertion of DNA templates in the correct position. The sequence of the bands was verified by Sanger sequencing.

For the *Tymp* locus, the band corresponding to the NHEJ insertion was not visible in a simple PCR (Fig. 41a), but it was detectable with a nested PCR (Fig. 41b). This means that the gene editing efficiency was low. For the *Alb* locus, only the band corresponding to HDR was visible both with a simple and nested PCR (Fig. 41a,b), while the NHEJ band was only visible in the nested PCR (Fig. 41b). The HDR band was much more intense, which means that the gene editing by HDR was more efficient *in vitro* in Hepa1c1c7 cells.

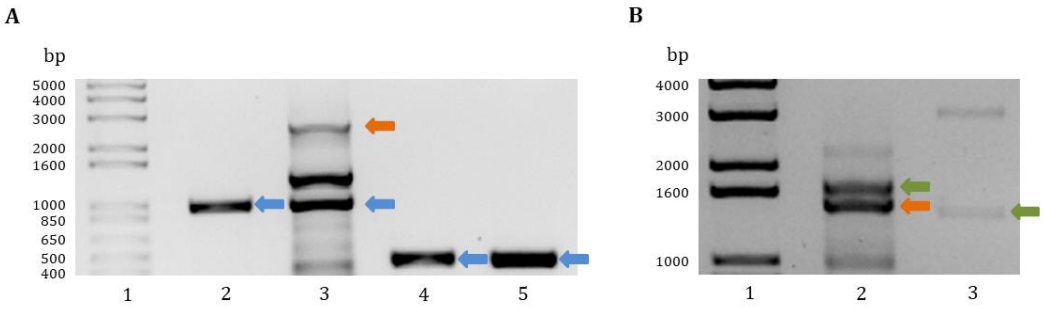

Figure 41. Detection of genomic insertion of the TP DNA templates. A) Representative image of simple PCR to detect gene editing. Lane 1: Size marker; lane 2: C- of *Alb* locus; lane 3: *Alb* locus gene editing sample; lane 4: C- of *Tymp* locus; lane 5: *Tymp* locus gene editing sample. **B)** Representative image of nested PCR to detect gene editing. Lane 1: Size marker; lane 2: *Alb* locus gene editing sample (transfected cells); lane 3: *Tymp* locus gene editing sample (transfected cells). Blue arrows point to unedited bands, green arrows point to NHEJ bands, and orange arrows point to HDR bands.

At the *Alb* locus, we relatively quantified the HDR gene editing though the intensity of the bands resulting from the simple PCR. For the first gene editing approach (CRISPR/Cas9 in RNA system), cells transfected with only the DNA template were used as internal (baseline) control. In this case, the gene editing efficiency was 13,88 ± 1,7 %. When cells were also transfected with the CRISPR/Cas9 system, the gene editing efficiency was a bit higher (16,7 ± 1,61 %), although the difference was not statistically significant. However, this condition was significantly different from the negative control (non-transfected cells, 0%). For the second gene editing approach (CRISPR/Cas9 in DNA system), the gene editing efficiency was 13,44 ± 0,71 % (Fig. 42).

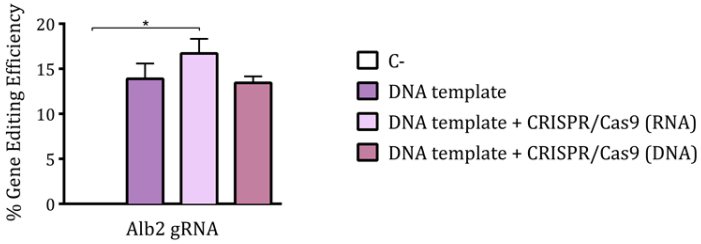

Figure 42. *In vitro* relative quantification of HDR gene editing in *Alb* locus. Error bars represent the SEM (n=3 in all groups). Data analysed with Kruskal-Wallis test with Dunn's multiple comparisons post hoc test. * = p<0,05.

5.1.3.2.2. hTP mRNA quantification (*Alb* locus)

At the *Alb* locus, we relatively quantified the hTP mRNA expression by real-time PCR (Fig. 43, Table 28). The expression observed after pCMV-Alb-TP plasmid transfection was significantly different from the negative control (non-transfected cells), which confirmed that the hybrid *Alb*-h*TP* gene was successfully transcribed in Hepa1c1c7 cells.

For the first gene editing approach (CRISPR/Cas9 in RNA system), cells transfected with only the DNA template had some basal hTP mRNA expression, but this value was higher when cells were also transfected with the CRISPR/Cas9 system, although the difference was not statistically significant. However, this condition was significantly different from the negative control. For the second gene editing approach (CRISPR/Cas9 in DNA system), there was some detectable hTP mRNA expression, but it was not statistically different form the negative control. Altogether, these results confirmed that edited cells successfully expressed Alb-hTP mRNA, even if the expression levels were very low.

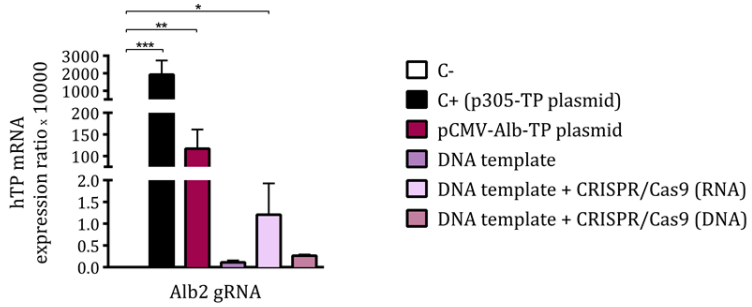

Figure 43. *In vitro* relative quantification of hTP mRNA. Error bars represent the SEM. Data analysed with Kruskal-Wallis test with Dunn's multiple comparisons post hoc test. * = p<0,05; ** = p<0,01; *** = p<0,001.

	C-	C+ (p305-TP plasmid)	pCMV-Alb-TP plasmid	DNA template	DNA template + CRISPR/Cas9 (RNA)	DNA template + CRISPR/Cas9 (DNA)
hTP mRNA expression ratio x 10000	0 (n=8)	1926 ± 808,3 (n=3)	117,1 ± 44,29 (n=3)	0,11 ± 0,03 (n=5)	1,21 ± 0,72 (n=8)	0,26 ± 0,03 (n=3)

Table 28. *In vitro* relative quantification of hTP mRNA. Values represent the hTP mRNA expression ratio (x 10000) ± SEM.

5.1.3.2.3. Alb-hTP expression

At the *Alb* locus, we assessed hTP protein expression by Western blot. This enzyme was undetectable in edited cell samples with both gene editing strategies. However, this technique confirmed that the Alb-hTP hybrid protein was successfully expressed in Hepa1c1c7 cells (Fig. 44a). The relative expression level observed after pCMV-Alb-TP

plasmid transfection was 29,99 ± 2,26 %, whereas the expression with p305-TP plasmid transfection was 100 ± 9,89 %. Although there was no statistical difference with the negative control (0%), the hybrid Alb-hTP protein was detectable and quantifiable (Fig. 44b).

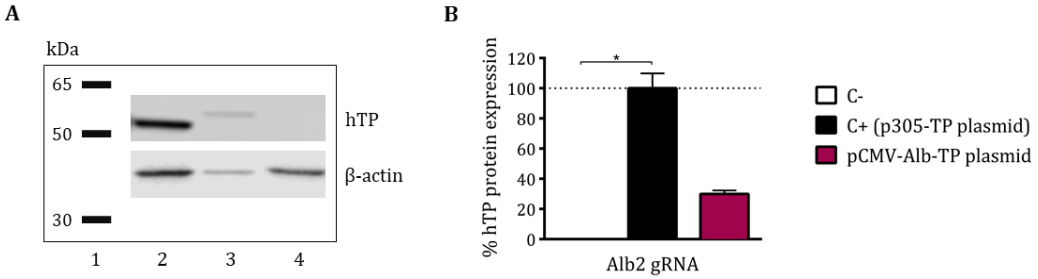

Figure 44. *In vitro* Alb-hTP protein detection and quantification. A) Representative image of a Western blot result. Lane 1: Size marker; lane 2: C+ (p305-TP plasmid transfection); lane 3: pCMV-Alb-TP plasmid transfection; lane 4: C- (non-transfected cells). **B)** Relative quantification of hTP protein expression. Error bars represent the SEM (n=3 in all groups). Data analysed with Kruskal-Wallis test with Dunn's multiple comparisons post hoc test. * = p<0,05; ** = p<0,01; *** = p<0,001.

5.1.3.2.4. Alb-hTP activity

To assess if the hybrid Alb-hTP protein maintained the enzymatic activity, we used Hepa1c1c7 cells transfected with the pCMV-Alb-TP plasmid. Lymphoblastoid cells, which were used as positive control, had an intrinsic activity of 522,57 ± 115,89 nmol Thy / h $_X$ mg protein. The intracellular basal activity level of Hepa1c1c7 cells was 10,27 ± 2,58 nmol Thy / h $_X$ mg protein, while these same cells transfected with the pCMV-Alb-TP plasmid had an activity of 36,5 ± 4,25 nmol Thy / h $_X$ mg protein (Fig. 45). Even if this difference was not statistically significant, we confirmed that the hybrid protein had some enzymatic activity. Additionally, since the hybrid protein could be secreted, TP activity was also assessed in the culture medium. In this case, medium from cells transfected with the pCMV-Alb-TP plasmid had an activity of 15,03 ± 4,79 nmol Thy / h $_X$ ml medium, while the basal extracellular activity was undetectable. This confirmed that the secreted Alb-hTP protein retained enzymatic activity (Fig. 45).

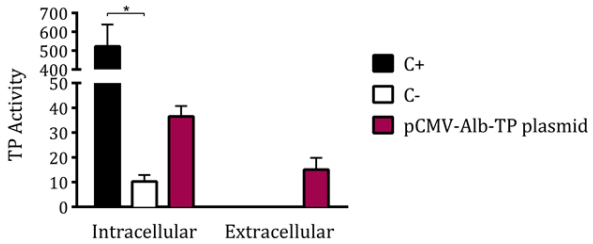

Figure 45. *In vitro* Alb-hTP hybrid protein activity. Intracellular TP activity units are nmol Thy / h $_X$ mg protein, while extracellular TP activity units are nmol Thy / h $_X$ ml medium. Error bars represent the SEM (n=3 in all groups). Data analysed with Kruskal-Wallis test with Dunn's multiple comparisons post hoc test. * = p<0,05.

5.2. *In vivo* experimentation

5.2.1. CRISPR/Cas9 efficiency

We tested the KH-ret formulation at different doses and weight ratios of gRNA : Cas9 mRNA. Although some animals presented a CRISPR/Cas9 cutting efficiency value higher than the background levels detected in negative control mice (injected with saline solution), the differences were not statistically significant in either loci (Fig. 46, Table 29).

We evaluated the LNP formulation in a single dose and weight ratio. In this case, the CRISPR/Cas9 cutting efficiency was clearly higher than background levels, although this difference was only statistically relevant in the *Tymp* locus (Fig. 46, Table 29).

Even if the differences were not statistically significant, it was clear that the LNP formulation was more efficient for *in vivo* liver gene editing.

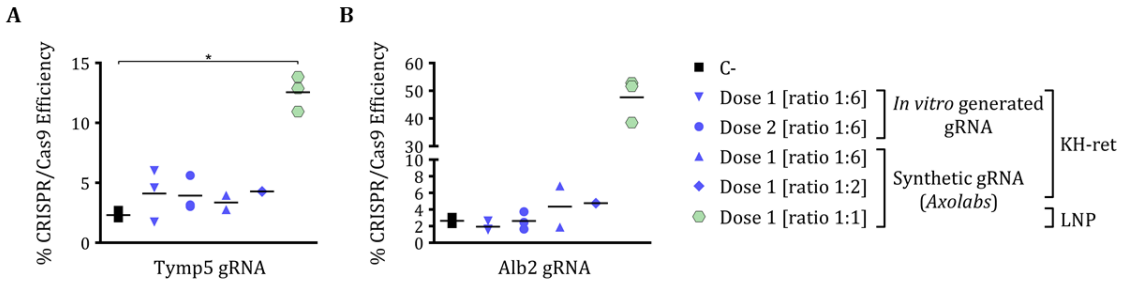

Figure 46. *In vivo* hepatic CRISPR/Cas9 cutting efficiency results measured by NGS. Percentage of CRISPR/Cas9 efficiency Tymp5 gRNA (**A**) and Alb2 gRNA (**B**). Each mark corresponds to a single mouse value and bars represent the mean of the group. Dose 1 corresponds to 50 µg RNA/mouse, while dose 2 corresponds to 100 µg RNA/mouse. The different [gRNA : Cas9 mRNA] weight ratios are specified. Data analysed with Kruskal-Wallis test with Dunn's multiple comparisons post hoc test. * = p<0,05.

			CRISPR/Cas9 efficiency (%)	
			Tymp5	**Alb2**
C-			2,31 ± 0,19 (n=3)	2,64 ± 0,22 (n=3)
KH-ret	**In vitro generated gRNAs**	**Dose 1 [ratio 1:6]**	4,11 ± 1,26 (n=3)	1,94 ± 0,34 (n=3)
		Dose 2 [ratio 1:6]	3,94 ± 0,84 (n=3)	2,61 ± 0,61 (n=3)
	Synthetic gRNAs (*Axolabs*)	**Dose 1 [ratio 1:6]**	3,36 ± 0,59 (n=2)	4,35 ± 2,47 (n=2)
		Dose 1 [ratio 1:2]	4,28 (n=1)	4,76 (n=1)
LNP		**Dose 1 [ratio 1:1]**	12,56 ± 0,85 (n=3)	47,66 ± 4,59 (n=3)

Table 29. *In vivo* hepatic CRISPR/Cas9 cutting efficiency results measured by NGS. Values represent the mean % ± SEM. Dose 1 corresponds to 50 µg RNA/mouse, while dose 2 corresponds to 100 µg RNA/mouse. The different [gRNA : Cas9 mRNA] weight ratios are specified.

5.2.2. Animal monitoring

5.2.2.1. Weight control

We monitored animal heath through the duration of experimentation. We weighed mice periodically to assess that their growth was correct. The weights of dKO treated mice did not differ from those of dKO untreated mice. This indicated that the treatments were probably not detrimental to their health. The growth rate of dKO mice was similar to that of WT mice. Both genders and gene editing strategies had a similar growth tendency (Fig. 47, 48).

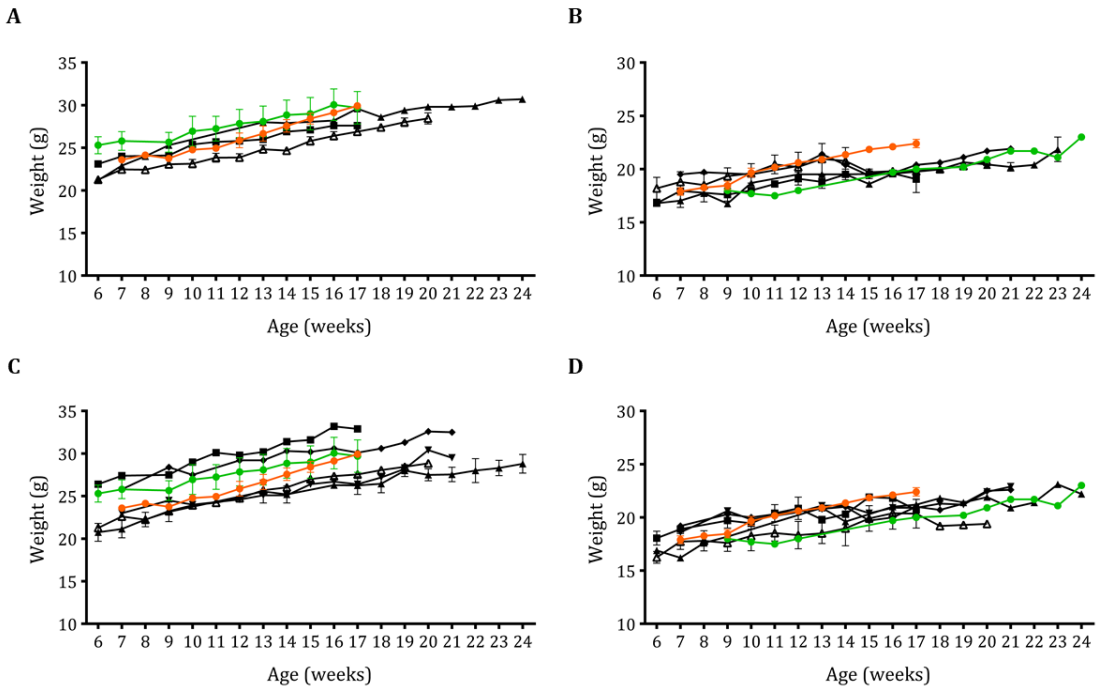

Figure 47. Weight monitoring: first gene editing approach (CRISPR/Cas9 delivered by NPs). **A)** *Tymp* locus gene editing in males; **B)** *Tymp* locus gene editing in females; **C)** *Alb* locus gene editing in males; **D)** *Alb* locus gene editing in females. Legend: ● = WT; ● = C- dKO; ■ = AAV template alone (10^{11} gc/mouse); ▲ = AAV template (10^{11} gc/mouse) + KH-ret (50 µg RNA/mouse [ratio 1:6]); ▼ = AAV template (2×10^{11} gc / mouse) + KH-ret (100 µg RNA/mouse [ratio 1:6]); ♦ = AAV template (10^{11} gc/mouse) + KH-ret (50 µg RNA/mouse [ratio 1:2]); △ = AAV template (4×10^{11} gc/mouse) + LNP (50 µg RNA/mouse [ratio 1:1]). Error bars represent the SEM.

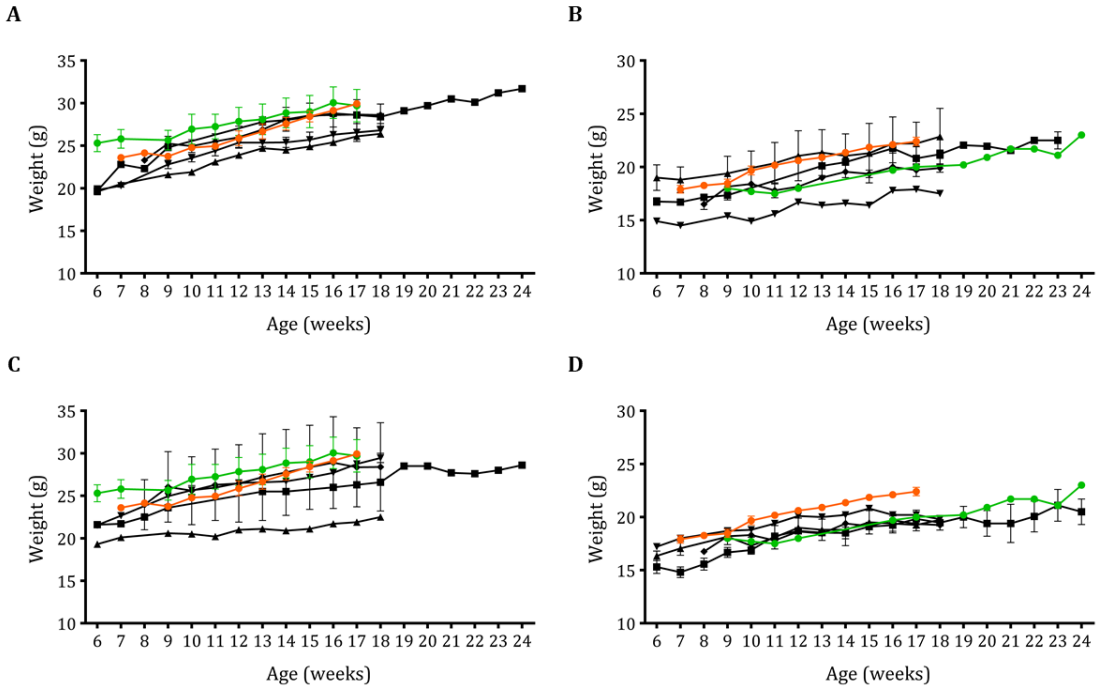

Figure 48. Weight monitoring: second gene editing approach (CRISPR/Cas9 delivered by AAV vectors). A) *Tymp* locus gene editing in males; **B)** *Tymp* locus gene editing in females; **C)** *Alb* locus gene editing in males; **D)** *Alb* locus gene editing in females. Legend: ● = WT; ● = C- dKO; ■ = Dose 1 (4×10^{10} gc/mouse of AAV-Cas9 + 10^{11} gc/mouse of AAV-gRNA-Template); ▲ = Dose 2 (8×10^{10} gc/mouse of AAV-Cas9 + 2×10^{11} gc/mouse of AAV-gRNA-Template); ▼ = Dose 3 ($1{,}6 \times 10^{11}$ gc/mouse of AAV-Cas9 + 4×10^{11} gc/mouse of AAV-gRNA-Template); ♦ = Dose 4 ($3{,}2 \times 10^{11}$ gc/mouse of AAV-Cas9 + 8×10^{11} gc/mouse of AAV-gRNA-Template). Error bars represent the SEM.

5.2.2.2. Thd and dUrd plasma levels

We quantified the Thd and dUrd basal plasma levels before treatment (6-9-week old animals). The levels in dKO mice were significantly higher than those of WT mice (Fig. 49, Table 30). In dKO mice, the basal plasma levels of both deoxyribonucleosides were significantly higher in females compared to males. This phenomenon was not observed in WT mice (Fig. 49, Table 30).

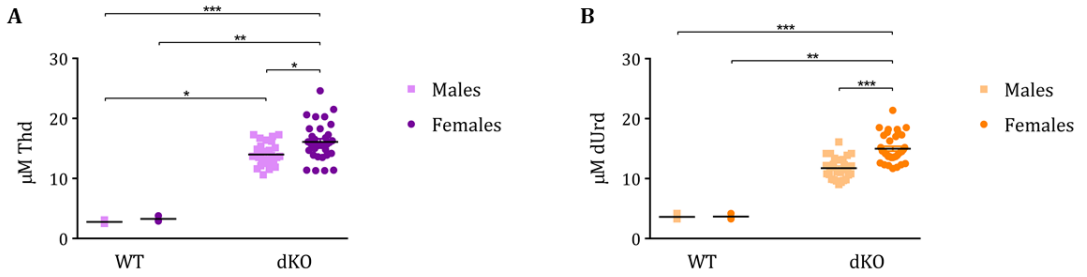

Figure 49. Thd and dUrd basal plasma levels. Quantification of plasma Thd (**A**) and dUrd (**B**). Each mark corresponds to a single mouse value and bars represent the mean of the group. Data analysed with Kruskal-Wallis test with Dunn's multiple comparisons post hoc test. * = p<0,05; ** = p<0,01; *** = p<0,001.

We measured the Thd and dUrd plasma levels after 2 and 4 weeks to assess whether the treatments were effective. The negative control group (injected with saline solution) and mice treated with the AAV templates alone showed no changes in Thd and dUrd levels through time.

In the first gene editing approach, we treated the animals with different doses and conditions of KH-ret PNP formulation and AAV templates. Although in some cases a slight decrease in levels was observed after treatment, there was no statistical difference with the pre-treatment values in both loci. However, mice treated with the LNP formulation and AAV templates had a statistically significant decrease in both deoxyribonucleosides levels after treatment, although these values did not reach the WT levels. Similar results were obtained in both loci (Fig. 50, Table 31).

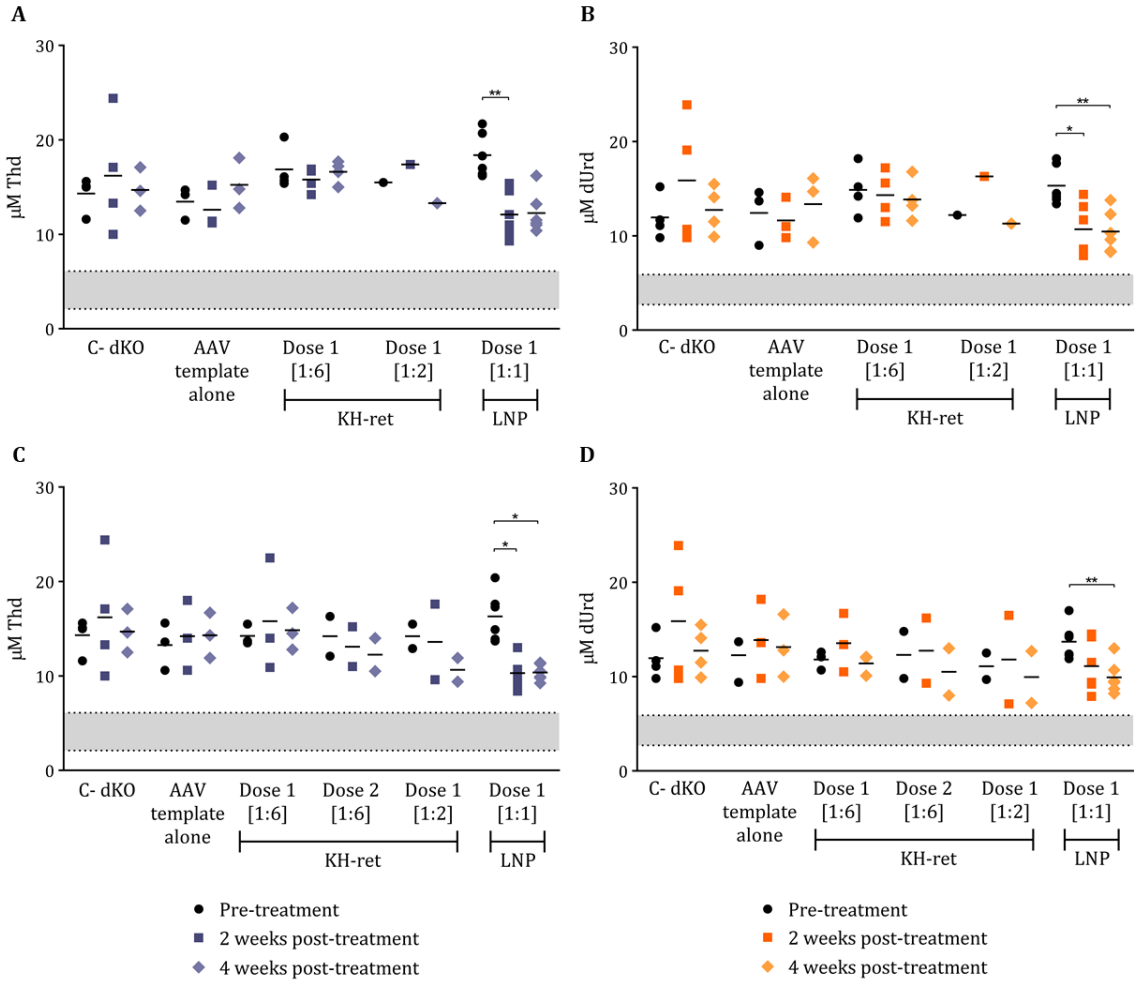

Figure 50. Screening of nucleoside plasma levels: first gene editing approach (CRISPR/Cas9 delivered by NPs). A) Thd levels in *Tymp* gene editing; **B)** dUrd levels in *Tymp* gene editing; **C)** Thd levels in *Alb* gene editing; **D)** dUrd levels in *Alb* gene editing. Experimental conditions: C- dKO; AAV template alone (10^{11} gc/mouse); KH-ret dose 1 (50 µg RNA/mouse) [ratio 1:6] + AAV template (10^{11} gc/mouse); KH-ret dose 2 (100 µg RNA/mouse) [ratio 1:6] + AAV template (2×10^{11} gc / mouse); KH-ret dose 1 (50 µg RNA/mouse) [ratio 1:2] + AAV template (10^{11} gc / mouse); LNP dose 1 (50 µg RNA/mouse) [ratio 1:1] + AAV template (4×10^{11} gc / mouse). The grey area represents the range found in WT mice. Each mark corresponds to a single mouse value and bars represent the mean of the group. Data analysed with Friedman test with Dunn's multiple comparisons post hoc test. * = p<0,05; ** = p<0,01.

In the second gene editing approach, we tested four different doses. Some slight decrease was observed in post-treatment values, but only in one case this decrease was statistically significant (dUrd levels in *Tymp* locus dose 2) (Fig. 51, Table 31).

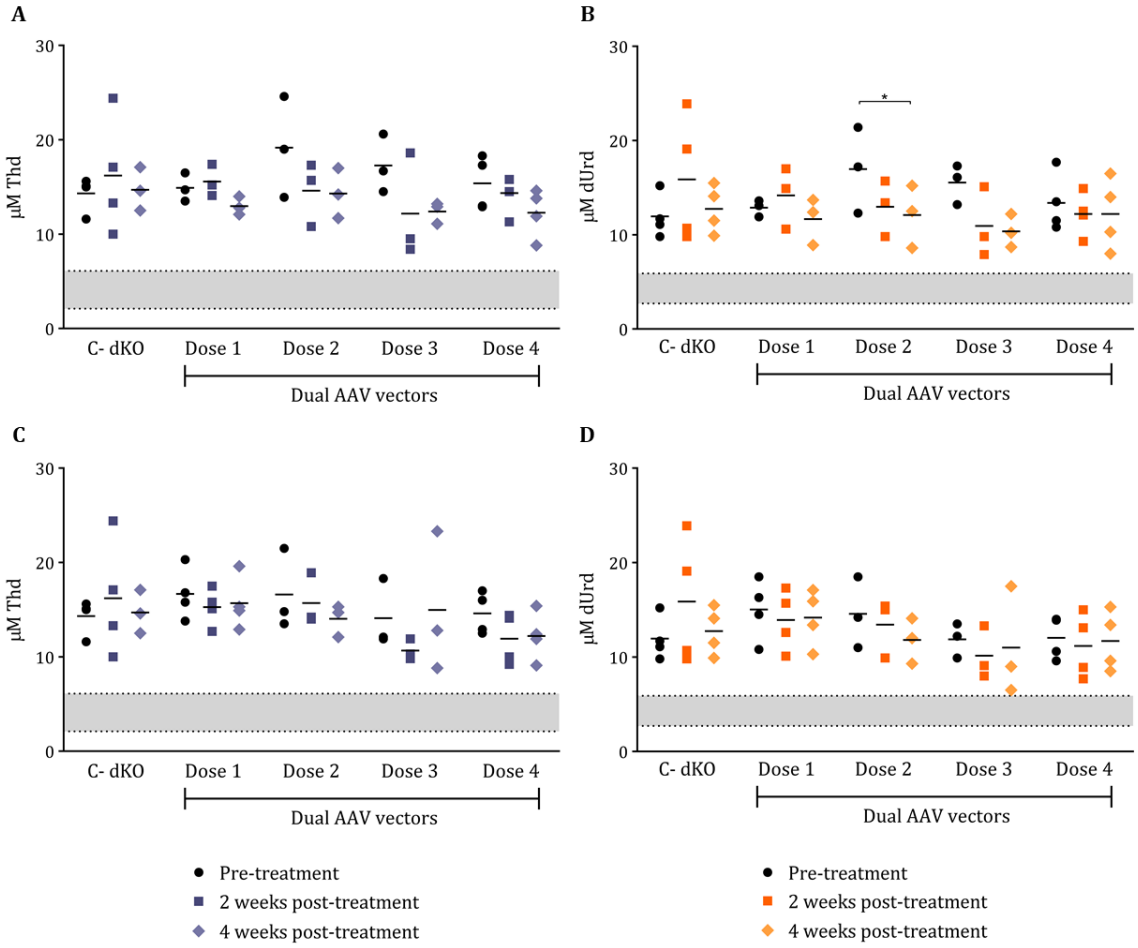

Figure 51. Screening of nucleoside plasma levels: second gene editing approach (CRISPR/Cas9 delivered by AAV vectors). A) Thd levels in *Tymp* gene editing; **B)** dUrd levels in *Tymp* gene editing; **C)** Thd levels in *Alb* gene editing; **D)** dUrd levels in *Alb* gene editing. Experimental conditions: C- dKO; Dose 1 (4×10^{10} gc/mouse of AAV-Cas9 + 10^{11} gc/mouse of AAV-gRNA-Template); Dose 2 (8×10^{10} gc/mouse of AAV-Cas9 + 2×10^{11} gc/mouse of AAV-gRNA-Template); Dose 3 ($1{,}6 \times 10^{11}$ gc/mouse of AAV-Cas9 + 4×10^{11} gc/mouse of AAV-gRNA-Template); Dose 4 ($3{,}2 \times 10^{11}$ gc/mouse of AAV-Cas9 + 8×10^{11} gc/mouse of AAV-gRNA-Template). The grey area represents the range found in WT mice. Each mark corresponds to a single mouse value and bars represent the mean of the group. Data analysed with Friedman test with Dunn's multiple comparisons post hoc test. * = $p<0{,}05$.

		Thd (µM)	dUrd (µM)
WT	**Males**	2,76 ± 0,11 (n=5)	3,62 ± 0,17 (n=5)
	Females	3,27 ± 0,27 (n=3)	3,67 ± 0,27 (n=3)
dKO	**Males**	13,98 ± 0,34 (n=29)	11,75 ± 0,33 (n=29)
	Females	16,09 ± 0,52 (n=34)	15 ± 0,4 (n=34)

Table 30. Thd and dUrd basal plasma levels. Values represent the plasma Thd and dUrd concentration (µM) ± SEM.

			Plasma nucleoside levels (Thd and dUrd)		
			Pre-treatment	Post-treatment	
				2 weeks	4 weeks
C- (WT)			**2,95 ± 0,14** **3,64 ± 0,13** (n=8)	**3,55 ± 0,44** **3,62 ± 0,36** (n=8)	**3,16 ± 0,26** **3,52 ± 0,23** (n=8)
C- (dKO)			**14,32 ± 0,92** **11,95 ± 1,15** (n=4)	**16,2 ± 3,09** **15,87 ± 3,4** (n=4)	**14,7 ± 0,94** **12,75 ± 1,26** (n=4)
Gene editing approach 1: NPs + rAAV2/8 vectors	***Tymp***	**AAV template alone (10^{11} gc/mouse)**	**13,47 ± 0,99** **12,43 ± 1,73** (n=3)	**12,6 ± 1,3** **11,63 ± 1,28** (n=3)	**15,23 ± 1,54** **13,37 ± 2,07** (n=3)
		AAV template (10^{11} gc/mouse) + KH-ret (50 µg RNA/mouse [ratio 1:6])	**16,87 ± 1,15** **14,87 ± 1,31** (n=4)	**15,8 ± 0,63** **14,32 ± 1,28** (n=4)	**16,62 ± 0,59** **13,85 ± 1,09** (n=4)
		AAV template (10^{11} gc/mouse) + KH-ret (50 µg RNA/mouse [ratio 1:2])	**15,5** **12,2** (n=1)	**17,4** **16,3** (n=1)	**13,3** **11,3** (n=1)
		AAV template (4×10^{11} gc/mouse) + LNP (50 µg RNA/mouse [ratio 1:1])	**18,38 ± 0,95** **15,31 ± 0,85** (n=6)	**12,1 ± 0,99** **10,7 ± 1,12** (n=6)	**12,25 ± 0,88** **10,45 ± 0,9** (n=6)
	Alb	**AAV template alone (10^{11} gc/mouse)**	**13,27 ± 1,45** **12,27 ± 1,43** (n=3)	**14,2 ± 2,14** **13,87 ± 2,43** (n=3)	**14,3 ± 1,39** **13,13 ± 1,91** (n=3)
		AAV template (10^{11} gc/mouse) + KH-ret (50 µg RNA/mouse [ratio 1:6])	**14,23 ± 0,64** **11,8 ± 0,57** (n=3)	**15,8 ± 3,47** **13,53 ± 1,79** (n=3)	**14,83 ± 1,28** **11,4 ± 0,65** (n=3)
		AAV template (2×10^{11} gc/mouse) + KH-ret (100 µg RNA/mouse [ratio 1:6])	**14,2 ± 2,1** **12,3 ± 2,5** (n=2)	**13,1 ± 2,1** **12,75 ± 3,45** (n=2)	**12,25 ± 1,75** **10,5 ± 2,5** (n=2)

		AAV template (10^{11} gc/mouse) + KH-ret (50 µg RNA/mouse [ratio 1:2])	**14,2 ± 1,3** **11,1 ± 1,4** (n=2)	**13,6 ± 4** **11,8 ± 4,7** (n=2)	**10,65 ± 1,25** **9,95 ± 2,75** (n=2)
		AAV template (4 x 10^{11} gc/mouse) + LNP (50 µg RNA/mouse [ratio 1:1])	**16,3 ± 1,06** **13,7 ± 0,79** (n=6)	**10,29 ± 0,63** **11,12 ± 1,13** (n=6)	**10,36 ± 0,35** **9,92 ± 0,71** (n=6)
Gene editing approach 2: dual rAAV2/8 vectors	***Tymp***	**Dose 1 (4 x 10^{10} gc/mouse of AAV-Cas9 + 10^{11} gc/mouse of AAV-gRNA-Template)**	**14,9 ± 0,87** **12,87 ± 0,5** (n=3)	**15,57 ± 0,97** **14,17 ± 1,88** (n=3)	**12,97 ± 0,55** **11,67 ± 1,43** (n=3)
		Dose 2 (8 x 10^{10} gc/mouse of AAV-Cas9 + 2 x 10^{11} gc/mouse of AAV-gRNA-Template)	**19,17 ± 3,09** **16,97 ± 2,63** (n=3)	**14,6 ± 1,95** **12,97 ± 1,71** (n=3)	**14,3 ± 1,53** **12,1 ± 1,91** (n=3)
		Dose 3 (1,6 x 10^{11} gc/mouse of AAV-Cas9 + 4 x 10^{11} gc/mouse of AAV-gRNA-Template)	**17,27 ± 1,78** **15,53 ± 1,21** (n=3)	**12,17 ± 3,23** **10,93 ± 2,15** (n=3)	**12,4 ± 0,66** **10,37 ± 1,01** (n=3)
		Dose 4 (3,2 x 10^{11} gc/mouse of AAV-Cas9 + 8 x 10^{11} gc/mouse of AAV-gRNA-Template)	**15,37 ± 1,41** **13,37 ± 1,55** (n=4)	**14,35 ± 1,06** **12,2 ± 1,15** (n=4)	**12,27 ± 1,29** **12,2 ± 1,89** (n=4)
	Alb	**Dose 1 (4 x 10^{10} gc/mouse of AAV-Cas9 + 10^{11} gc/mouse of AAV-gRNA-Template)**	**16,67 ± 1,36** **15,02 ± 1,63** (n=4)	**15,27 ± 0,99** **13,92 ± 1,6** (n=4)	**15,67 ± 1,41** **14,17 ± 1,5** (n=4)
		Dose 2 (8 x 10^{10} gc/mouse of AAV-Cas9 + 2 x 10^{11} gc/mouse of AAV-gRNA-Template)	**16,6 ± 2,48** **14,57 ± 2,17** (n=3)	**15,7 ± 1,6** **13,43 ± 1,77** (n=3)	**14,03 ± 0,98** **11,8 ± 1,39** (n=3)
		Dose 3 (1,6 x 10^{11} gc/mouse of AAV-Cas9 + 4 x 10^{11} gc/mouse of AAV-gRNA-Template)	**14,1 ± 2,1** **11,87 ± 1,05** (n=3)	**10,67 ± 0,63** **10,13 ± 1,61** (n=3)	**14,97 ± 4,32** **11 ± 3,33** (n=3)
		Dose 4 (3,2 x 10^{11} gc/mouse of AAV-Cas9 + 8 x 10^{11} gc/mouse of AAV-gRNA-Template)	**14,6 ± 1,12** **12,02 ± 1,13** (n=4)	**11,92 ± 1,35** **11,17 ± 1,72** (n=4)	**12,2 ± 1,29** **11,7 ± 1,59** (n=4)

Table 31. Screening of nucleoside plasma levels. Values represent the plasma nucleoside concentration (µM) ± SEM. Thd in violet and dUrd in orange. The different [gRNA : Cas9 mRNA] weight ratios were specified.

5.2.3. Hepatic gene editing assessment

5.2.3.1. Genomic insertion

We assessed the insertion of DNA templates in the correct loci by PCR using specific primers (livers extracted 5-10 weeks after injection). In some cases the insertion was detectable (Table 32). The insertion through HDR was easier to detect as the size of the resulting PCR product was more stable. On the other hand, the detection of NHEJ-directed insertion was harder to define, as the size of the resulting PCR product could vary.

In the first gene editing approach, we observed the best results in mice injected with the LNP formulation and AAV templates. In the second gene editing approach, we achieved the best results in mice injected with the highest doses (Table 32).

			Proportion of animals with detectable insertion
C- (WT)			0/3
C- (dKO)			0/3
Gene editing approach 1: NPs + rAAV2/8 vectors	***Tymp***	**AAV template alone (10^{11} gc/mouse)**	0/3
		AAV template (10^{11} gc/mouse) + KH-ret (50 µg RNA/mouse [ratio 1:6])	0/4
		AAV template (10^{11} gc/mouse) + KH-ret (50 µg RNA/mouse [ratio 1:2])	0/1
		AAV template (4×10^{11} gc/mouse) + LNP (50 µg RNA/mouse [ratio 1:1])	3/3
	Alb	**AAV template alone (10^{11} gc/mouse)**	1/3
		AAV template (10^{11} gc/mouse) + KH-ret (50 µg RNA/mouse [ratio 1:6])	0/3
		AAV template (2×10^{11} gc/mouse) + KH-ret (100 µg RNA/mouse [ratio 1:6])	1/2
		AAV template (10^{11} gc/mouse) + KH-ret (50 µg RNA/mouse [ratio 1:2])	0/2
		AAV template (4×10^{11} gc/mouse) + LNP (50 µg RNA/mouse [ratio 1:1])	3/3
Gene editing approach 2: dual rAAV2/8 vectors	***Tymp***	**Dose 1 (4×10^{10} gc/mouse of AAV-Cas9 + 10^{11} gc/mouse of AAV-gRNA-Template)**	0/3
		Dose 2 (8×10^{10} gc/mouse of AAV-Cas9 + 2×10^{11} gc/mouse of AAV-gRNA-Template)	0/3
		Dose 3 ($1{,}6 \times 10^{11}$ gc/mouse of AAV-Cas9 + 4×10^{11} gc/mouse of AAV-gRNA-Template)	0/3

		Dose 4 (3,2 x 10^{11} gc/mouse of AAV-Cas9 + 8 x 10^{11} gc/mouse of AAV-gRNA-Template)	3/4
	Alb	**Dose 1 (4 x 10^{10} gc/mouse of AAV-Cas9 + 10^{11} gc/mouse of AAV-gRNA-Template)**	0/4
		Dose 2 (8 x 10^{10} gc/mouse of AAV-Cas9 + 2 x 10^{11} gc/mouse of AAV-gRNA-Template)	2/3
		Dose 3 (1,6 x 10^{11} gc/mouse of AAV-Cas9 + 4 x 10^{11} gc/mouse of AAV-gRNA-Template)	1/3
		Dose 4 (3,2 x 10^{11} gc/mouse of AAV-Cas9 + 8 x 10^{11} gc/mouse of AAV-gRNA-Template)	2/4

Table 32. Hepatic genomic insertion of the templates. Values represent the proportion of animals that presented detectable insertion in each group.

5.2.3.2. hTP mRNA quantification (*Alb* locus)

We relatively quantified the hTP mRNA expression in the RNA extracts of livers of *Alb* locus edited mice (5-10 weeks after injection). The negative control group presented undetectable levels, while all groups of treated mice presented some detectable hTP mRNA expression.

In the first gene editing approach, mice treated with different conditions of KH-ret PNP formulation and AAV templates presented very low levels of hTP mRNA expression. Even animals treated with the AAV templates alone presented higher levels. However, the hTP mRNA expression levels observed in mice treated with the LNP formulation and AAV templates were much higher and significantly different from the negative control group (Fig. 52, Table 33).

In the second gene editing approach, the hTP mRNA expression levels presented a dose-dependent increase. However, only dose 4 was significantly different from the negative control group (Fig. 52, Table 33).

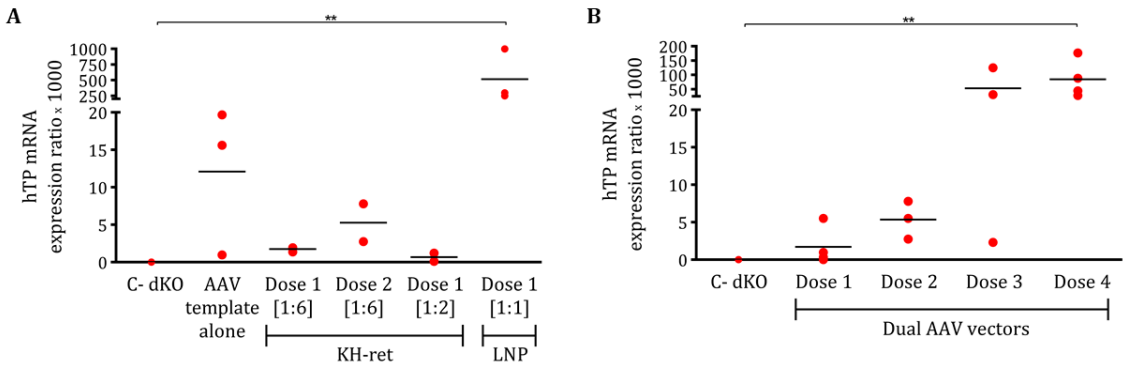

Figure 52. *In vivo* relative quantification of liver hTP mRNA. A) First gene editing approach (CRISPR/Cas9 delivered by NPs); **B)** Second gene editing approach (CRISPR/Cas9 delivered by AAV vectors). Experimental conditions: C- dKO; AAV template alone (10^{11} gc/mouse); KH-ret dose 1 (50

µg RNA/mouse) [ratio 1:6] + AAV template (10^{11} gc/mouse); KH-ret dose 2 (100 µg RNA/mouse) [ratio 1:6] + AAV template (2×10^{11} gc / mouse); KH-ret dose 1 (50 µg RNA/mouse) [ratio 1:2] + AAV template (10^{11} gc / mouse); LNP dose 1 (50 µg RNA/mouse) [ratio 1:1] + AAV template (4×10^{11} gc / mouse); Dual AAV dose 1 (4×10^{10} gc/mouse of AAV-Cas9 + 10^{11} gc/mouse of AAV-gRNA-Template); Dual AAV dose 2 (8×10^{10} gc/mouse of AAV-Cas9 + 2×10^{11} gc/mouse of AAV-gRNA-Template); Dual AAV dose 3 ($1,6 \times 10^{11}$ gc/mouse of AAV-Cas9 + 4×10^{11} gc/mouse of AAV-gRNA-Template); Dual AAV dose 4 ($3,2 \times 10^{11}$ gc/mouse of AAV-Cas9 + 8×10^{11} gc/mouse of AAV-gRNA-Template). Each mark corresponds to a single mouse value and bars represent the mean of the group. Data analysed with Kruskal-Wallis test with Dunn's multiple comparisons post hoc test. ** = p<0,01.

		hTP mRNA expression ratio x 1000
C- (dKO)		0 (n=4)
Gene editing approach 1: NPs + rAAV2/8 vectors	**AAV template alone (10^{11} gc/mouse)**	12,08 ± 5,68 (n=3)
	AAV template (10^{11} gc/mouse) + KH-ret (50 µg RNA/mouse [ratio 1:6])	1,75 ± 0,19 (n=3)
	AAV template (2×10^{11} gc/mouse) + KH-ret (100 µg RNA/mouse [ratio 1:6])	5,27 ± 2,52 (n=2)
	AAV template (10^{11} gc/mouse) + KH-ret (50 µg RNA/mouse [ratio 1:2])	0,66 ± 0,55 (n=2)
	AAV template (4×10^{11} gc/mouse) + LNP (50 µg RNA/mouse [ratio 1:1])	515,8 ± 242,5 (n=3)
Gene editing approach 2: dual rAAV2/8 vectors	**Dose 1 (4×10^{10} gc/mouse of AAV-Cas9 + 10^{11} gc/mouse of AAV-gRNA-Template)**	1,7 ± 1,28 (n=4)
	Dose 2 (8×10^{10} gc/mouse of AAV-Cas9 + 2×10^{11} gc/mouse of AAV-gRNA-Template)	5,35 ± 1,46 (n=3)
	Dose 3 ($1,6 \times 10^{11}$ gc/mouse of AAV-Cas9 + 4×10^{11} gc/mouse of AAV-gRNA-Template)	52,84 ± 37,03 (n=3)
	Dose 4 ($3,2 \times 10^{11}$ gc/mouse of AAV-Cas9 + 8×10^{11} gc/mouse of AAV-gRNA-Template)	84,28 ± 33,37 (n=4)

Table 33. *In vivo* relative quantification of liver hTP mRNA. Values represent the hTP mRNA expression ratio ($\times 1000$) ± SEM. The different [gRNA : Cas9 mRNA] weight ratios were specified.

5.2.3.3. TP expression

We analysed the TP expression in the protein extracts of livers (5-10 weeks after injection). This technique did not allow the detection of TP in dKO treated mice, except in animals injected with the LNP formulation and AAV templates.

TP expression in WT mice was 100 ± 15,38 %, while it was undetectable in dKO untreated mice. For the *Tymp* locus, TP expression was 7,02 ± 1,46 %, whereas for the *Alb* locus it was 6,45 ± 0,7 % (Fig. 53).

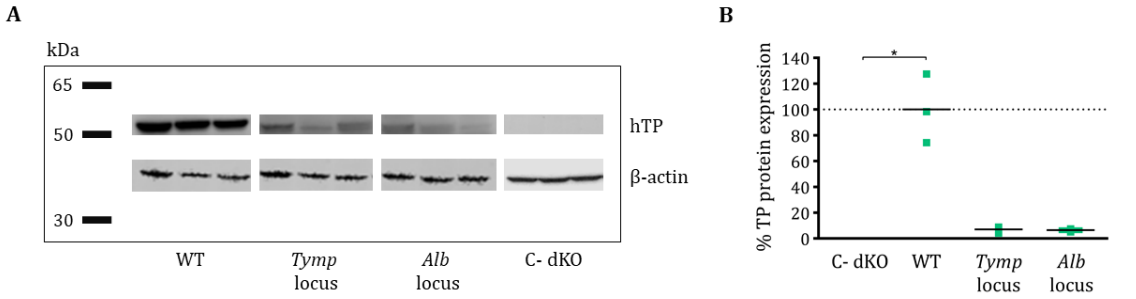

Figure 53. *In vivo* liver TP protein detection and quantification: mice treated with LNP formulation and AAV templates. A) Western blot results. **B)** Relative quantification of TP protein expression. Animals injected with LNP dose 1 (50 µg RNA/mouse) [ratio 1:1] + AAV template (4 x 10^{11} gc / mouse). Each mark corresponds to a single mouse value and bars represent the mean of the group (n=3 in all groups). Data analysed with Kruskal-Wallis test with Dunn's multiple comparisons post hoc test. * = p<0,05.

5.2.3.4. TP activity

We quantified the intracellular TP activity in the protein extracts of livers (5-10 weeks after injection). In both loci, the enzymatic activity values were very low compared to the WT values. The highest activity was detected in mice treated with LNPs and AAV vectors (Fig. 54, Table 34). We also assessed the TP activity in plasma samples of some *Alb* locus edited mice, as the resulting enzyme was secreted. In this case, mice treated with LNPs and AAV vectors presented an activity of 133,8 ± 34,67 nmol Thy / h x ml plasma. Mice treated with dose 4 of dual AAV vectors had undetectable plasma activity (Fig. 54b).

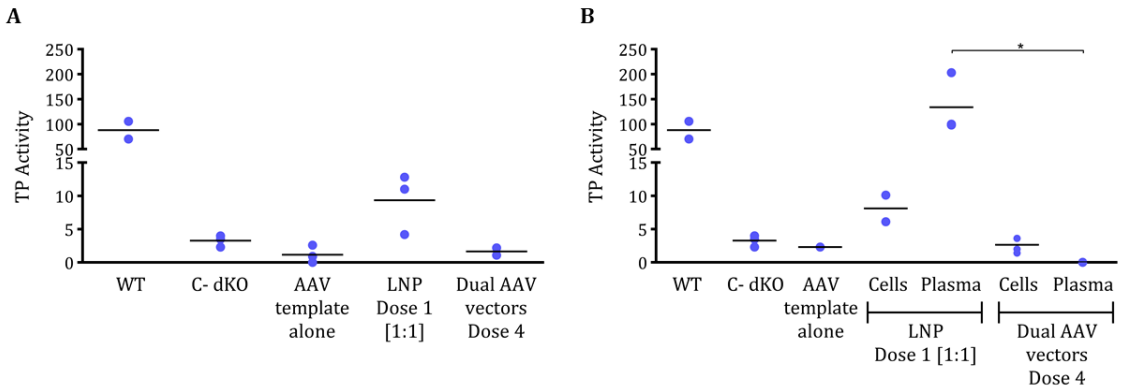

Figure 54. *In vivo* TP activity. A) *Tymp* gene editing; **B)** *Alb* gene editing. Experimental conditions: WT; C- dKO; AAV template alone (10^{11} gc/mouse); LNP dose 1 (50 µg RNA/mouse) [ratio 1:1] + AAV template (4 x 10^{11} gc / mouse); Dual AAV dose 4 (3,2 x 10^{11} gc/mouse of AAV-Cas9 + 8 x 10^{11} gc/mouse of AAV-gRNA-Template). Intracellular TP activity units are nmol Thy / h x mg protein, while plasma TP activity units are nmol Thy / h x ml plasma. Error bars represent the SEM. Each mark corresponds to a single mouse value and bars represent the mean of the group. Data analysed with Kruskal-Wallis test with Dunn's multiple comparisons post hoc test. * = p<0,05.

			TP activity
C- (WT)			87,85 ± 17,75 (n=2)
C- (dKO)			3,27 ± 0,5 (n=3)
Gene editing approach 1: NPs + rAAV2/8 vectors	***Tymp***	**AAV template alone (10^{11} gc/mouse)**	1,17 ± 0,76 (n=3)
		AAV template (4 x 10^{11} gc/mouse) + LNP (50 µg RNA/mouse [ratio 1:1])	9,33 ± 2,62 (n=3)
	Alb	**AAV template alone (10^{11} gc/mouse)**	2,3 (n=1)
		AAV template (4 x 10^{11} gc/mouse) + LNP (50 µg RNA/mouse [ratio 1:1])	8,1 ± 2 (n=2)
Gene editing approach 2: dual rAAV2/8 vectors	***Tymp***	**Dose 4 (3,2 x 10^{11} gc/mouse of AAV-Cas9 + 8 x 10^{11} gc/mouse of AAV-gRNA-Template)**	1,65 ± 0,78 (n=2)
	Alb	**Dose 4 (3,2 x 10^{11} gc/mouse of AAV-Cas9 + 8 x 10^{11} gc/mouse of AAV-gRNA-Template)**	2,65 ± 0,56 (n=4)

Table 34. *In vivo* liver TP activity. Values represent the TP protein activity (nmol Thy / h x mg protein) ± SEM.

6. DISCUSSION

MNGIE is rare progressive and degenerative disease caused by pathogenic mutations in the *TYMP* gene, which encodes the cytoplasmic protein TP [8,23,24,25]. The deficiency of this enzyme ultimately produces an imbalance in mitochondrial dNTP pools and has a mutagenic effect in mtDNA, which is deleterious for cells [3,13,29,39,40,41,42,43,44].

The current management is primarily supportive [2,59], although several therapeutic approaches have been proposed [29,60,63,64,65,66,67,68,69,70,71,72,73,74,75,76,77,78,79,80,81,81,82,83]. Gene therapy could be a good alternative for MNGIE treatment. In our institution, the group of Dr. R. Martí, in collaboration with our group, has evaluated both LV [77,88,314,315,316] and AAV vectors [314,317,318] in preclinical studies. Although LV vectors achieved a lifelong biochemical correction, this procedure required myelosuppression, which is extremely risky in MNGIE patients and is associated with high rates of mortality [70,71]. Moreover, since they are integrative vectors, there is a potential risk of insertional mutagenesis. On the other hand, AAV vectors achieved the decrease of Thd and dUrd plasma levels without significant toxicity, but this biochemical correction decreased in a follow-up study. Although the deoxyribonucleoside levels were still in the therapeutic range in animals treated with the highest dose, this progressive loss of effectivity could be a problem for clinical application. Mice are not the best model to study long term effects, as their life expectancy is not comparable to humans. Since AAV vectors are not integrative, this loss of effect was not unexpected [151,169]. Additionally, AAV vectors present other important limitations, like their limited packaging capacity [223,224] and the immunogenicity of the capsid proteins [151,196,243], which precludes subsequent re-administration of the same or similar vectors.

To tackle the problem of loss of efficiency at long term, we propose the use CRISPR/Cas9 in a hepatocyte-directed gene editing approach. The administration of this nuclease targeting specific loci could enhance the insertion of a DNA template carrying the *TYMP* cDNA. This template would be inserted downstream a genomic promoter, so the cDNA would only be expressed if a correct insertion was achieved. This way, the hepatocytes would become permanently corrected, and theoretically transgene expression would not be lost at long term. Two genomic loci were explored, the endogenous *Tymp* locus, and the *Alb* locus, which contains a strong promoter that would allow a potent cDNA expression in edited hepatocytes [312,343,344]. The initial idea was to deliver CRISPR/Cas9 RNA molecules with non-viral vectors (NPs), and the DNA templates with rAAV2/8 vectors, all of them targeting the liver. This way, the expression of the Cas9 nuclease would be strong but transient, which would decrease the risk of off-target effects and immunogenicity [454]. Such strategy would combine an efficient system to deliver the RNA payload to the cytoplasm, where it is required for translation, with viral vectors that are very good at delivering their genomes (including templates designed for DNA repair) to the nucleus. Several studies have reported the successful delivery of CRISPR/Cas9 RNAs with different NP formulations [413,436,478,479,480]. In this work, we tested two types of NPs: PNPs of GEMAT-IQS, and LNPs of Acuitas Therapeutics®. On the other hand, rAAV2/8 vectors were shown to be highly efficient at targeting mouse liver [214,215,216,217,218,219,220,221]. Later, we proposed the delivery of both CRISPR/Cas9 and templates in rAAV2/8 vectors as an alternative approach. Since the efficiency of the tested NPs at hepatocyte targeting was uncertain, this strategy served as a proof of concept to ascertain that the gene editing strategy was feasible. In this case, CRISPR/Cas9 expression would not be transient, so the risk of

immunogenicity and off-target effect would be higher, but we would be able to confirm if the templates were successfully inserted and expressed.

The liver has always been an attractive option as a target tissue, as it presents a highly fenestrated endothelium that favours the uptake of many molecules. Moreover, due to its anatomical characteristics, vectors administered by peripheral blood injection tend to accumulate in this tissue, which favours their uptake by hepatic cells [181,195]. Also, many studies have shown that hepatocyte-restricted transgene expression at high levels can induce transgene-specific immune tolerance [199,200,201,202,203,204,205,206,207]. For all these reasons, we hypothesized that hepatocytes would be the ideal target cells for our gene editing approach.

6.1. LNPs are more efficient in CRISPR/Cas9 hepatocyte delivery

The results of this work clearly demonstrate that, under our experimental conditions, LNPs from Acuitas Therapeutics® result in a higher hepatic gene editing efficiency than PNPs from GEMAT-IQS. These differences can be explained by multiple factors.

LNPs were already optimized for mRNA hepatic delivery [309,310,311,312], and the company had experience in encapsulating CRISPR/Cas9 RNA molecules in these particles. Moreover, several studies corroborate that LNPs in general have a natural affinity for liver tissue when administered systemically [304,305,413,501,502,503]. On the other hand, PNPs from GEMAT-IQS were not initially designed to target the liver [291,293,297]. The KH formulation was specifically modified by adding retinol molecules to target the liver. However, this KH-ret formulation preferentially transfected non-hepatocytic cells [294]. In our study, the gene editing had to be directed specifically to hepatocytes, as they are the only cell type with strong Alb (and probably also TP) expression. Altogether, the current KH-ret formulation was not the most adequate to target hepatocytes, and further optimization is needed.

Another factor that may have influenced the CRISPR/Cas9 efficiency results obtained with KH-ret formulation is its tendency to form aggregates. These PNPs were delivered in a lyophilized form, and the user had to reconstitute them immediately before administration. However, it was a delicate process that had to be performed rapidly and accurately. The lyophilized NPs had to be maintained at low temperature until just before they were resuspended. Slight increases in temperature caused the formation of aggregates upon resuspension. Moreover, if aggregates were formed, there was no way to reverse the situation. Also, the aggregates were not always visible, so it was difficult to determine if the batch had to be discarded. But the presence of aggregates was clearly visible upon administration to mice. In the majority of cases, the animals remained immobile after the injection, and it took hours before they started moving normally again. In some cases (20% approximately), animals died of cardiorespiratory arrest within minutes after injection. The surviving animals did not present any further clinical sign. However, it was clear that the compound administered caused some sort of acute toxicity to the animals. Indeed, this was a very limiting factor in the KH-ret treated groups. The presence of aggregates most likely reduced their cell transfection capacity, which is another factor that could explain their low efficiency in CRISPR/Cas9 delivery to the liver. This problem could be partly related to our lack of experience in PNP reconstitution, but it

is important to note that this phenomenon also happened when the resuspension was performed by an expert (a GEMAT-IQS scientist skilled in KH-ret reconstitution), although in this case no animals died. Nonetheless, the irreversibility of aggregates formation is a problem that should be tackled in the future by our GEMAT-IQS collaborators. On the other hand, LNPs were delivered already in suspension and did not form aggregates. No animal died after injection, and mice did not present any apparent signs of toxicity.

The weight ratio gRNA : mRNA Cas9 used in both NP formulations was different, which could also contribute to their efficiency differences. The first NPs we tested were the PNPs from GEMAT-IQS. Since we obtained good results *in vitro* with the 1:6 ratio, we decided to use this one during *in vivo* experimentation. Later, we contacted Acuitas Therapeutics® to test their LNP formulation under a scientific collaboration, and they recommended us to use a 1:1 ratio, as they had previously observed better results with it. Following their advice and some results published [413], we decided to switch ratios. By that time, we were almost done with KH-ret experiments, but in the last experiment we tested the 1:2 ratio (1:1 ratio was not possible due to limiting gRNA quantity). It is important to note that we did not observe significant differences between the 1:6 and the 1:2 ratios, so this was most likely not the main cause of KH-ret relatively poor efficiency.

The high efficiency of LNPs in liver targeting raises the question of the risks of off-target effects caused by CRISPR/Cas9. It is known that this nuclease has tolerance for sequence mismatch, so it can cause off-target cleavage that could potentially lead to genotoxicity, genome instability, gene function disruption, epigenetic alterations or carcinogenesis [388,389,390]. CRISPR/Cas9 specificity is influenced by several factors, like the intrinsic nature of Cas9 or the structure and composition of gRNAs. But the abundance of effective CRISPR/Cas9 complexes relative to the genomic target concentration is also an important factor [329]. For this reason, the high efficiency of LNPs could potentially cause an increase in off-target effects. The concentration of CRISPR/Cas9 delivered to cells should be adjusted to the minimum necessary for efficient cleavage, as the presence of surplus molecules increases the chance of off-target cleavage [329,388,397]. One of the reasons of delivering CRISPR/Cas9 as RNA molecules is to lessen the risk of off-target mutations, as their presence inside cells is transient. However, in this work the off-target effects were not evaluated. Although the gRNAs used had chemical modifications to enhance stability and efficacy (Fig. 11) [413], their off-target effects should have to be determined empirically. In the future we intend to study whether the two selected gRNAs (Tymp5 and Alb2) cause a significant off-target mutagenesis in the livers of mice treated with LNPs and AAV vectors. Moreover, the CRISPR/Cas9 dose should also have to be empirically determined by maximizing their efficiency while minimizing the off-target risk. In conclusion, the study of off-target effects is an important point that needs to be taken into account if this gene editing approach reaches clinical development in the future.

6.2. Gene editing with LNPs was more efficient than with dual rAAV2/8 vectors

Although we initially developed the strategy of delivering both CRISPR/Cas9 and templates in rAAV2/8 vectors as a way to assess that the gene editing approach we

designed was feasible, we found that the dual AAV approach presented worse results than that using CRISPR/Cas9 delivered as RNAs in LNPs.

We think that the timing of the injections could be one of the determining factors. Due to time limitations, the Thd and dUrd levels were only screened during a month post-treatment, and the livers were extracted soon after (although a cohort of LNP-treated mice is still being monitored for a long-term follow up). It is known that AAV vectors have a delayed onset of gene expression [100,165,166], so maybe a month was not enough to reach sufficient levels of CRISPR/Cas9 active molecules. On the other hand, CRISPR/Cas9 RNA molecules delivered by LNPs are expressed during the first 24 hours, with the peak being approximately at 4 hours after administration [309,312]. The templates are also available early inside cells as ssDNA molecules. The delayed onset expression in AAV vectors does not apply in this case, as the templates do not need to be expressed; they just need to be present in the nucleus. In future experiments it would be interesting to follow-up those mice during a longer time to determine if the dual AAV gene editing approach has a delayed efficiency.

Another factor that may explain the differences in efficiencies between both approaches is the intrinsic structure of the AAV genomes used. While the composition of the templates was identical, the AAV genomes that contained them were not the same. When CRISPR/Cas9 was delivered in NPs, the AAV genome only contained the DNA templates (plus stuffer DNA downstream and the ITRs). On the other hand, in the dual AAV approach, the DNA templates shared space with the gRNA gene. This change was necessary due to AAV genome packaging limitation [223,224]. The large size of Cas9 gene required the packaging of Cas9 and gRNA in separate AAV vectors. This same strategy has already been successfully tested in other studies [396,424,435,437,458,459,460]. So, in theory, this change was not expected to affect the template insertion efficiency, but it is a difference that must be taken into account when comparing results.

Altogether, it is important to note that both strategies cannot be directly compared, as there are many differential factors that might have influenced the results.

6.3. Gene editing with LNPs resulted in consistent biochemical correction

Among all the conditions tested *in vivo*, the delivery of CRISPR/Cas9 with LNPs was by far the one with the best results. These animals were the only ones that attained a robust and consistent reduction of Thd and dUrd plasma levels. Moreover, these mice presented the highest levels of TP mRNA, and were the only animals in which the TP protein was detectable by Western blot. Their TP enzymatic activity was also the highest. Overall, it can be concluded that this group of animals presented a substantial amount of successful gene editing, and that the inserted cDNAs were able to be correctly transcribed and translated into functional TP molecules.

However, it is important to note that the biochemical correction obtained did not reach WT levels. Although both Thd and dUrd levels decreased significantly (Fig. 50, 51), they were still quite far from the normal levels found in WT mice. It is hard to know whether

such a reduction would have a therapeutic effect in clinical application, as this animal model only presents the biochemical characteristics of MNGIE patients. Indeed, this is one of the main limitations of the model used, as mice do not present any clinical manifestation. This difference between species can be explained by several factors. Mice present a shorter lifespan, so it is possible that they do not live long enough to accumulate sufficient mtDNA alterations in most tissues. Another factor may be that the dNTP imbalance in less dramatic in mice, so the impact on mtDNA is lower. Other intrinsic differences between species may also play a role in this phenomenon [42,44,88,89]. On the other hand, the fact that mice do not present the disease phenotype has its advantages. Their handling is easier and, since they have a normal lifespan, it is possible to follow-up treated animals at long-term. Altogether, maybe mice are not the most suited animal model for MNGIE, but they are useful to detect biochemical correction, which is not only a major feature of the disease, but it is also clearly pathogenic in humans. Due to this model limitation, we can speculate about whether the nucleotide level decrease observed in this work would translate into some kind of clinical improvement. In the future it would be interesting to develop another model that better reproduces the clinical features of MNGIE.

Another factor that must be taken into account is the fact that the basal Thd and dUrd plasma levels in dKO mice were higher than those observed in previous studies [313,314,316,317,318]. It is important to note that the animals used in this study are descendants of the same dKO colony used in these same studies. However, the housing facilities were different, as a new building was recently constructed. Not only the housing room was different, but the cages and diet were also modified. One study suggested that diet, host genetics and gut microbiota interact to create distinct responses in plasma metabolites [504]. Another work also concluded that stress disturbed the metabolome of blood plasma and urine of diabetic rats [505]. These studies demonstrate that the change of housing conditions could have had an impact on the nucleoside plasma levels of dKO mice. Interestingly, the Thd and dUrd plasma levels of WT mice did not change, although their living conditions were changed in the same manner. Perhaps WT mice have enough overall TP activity (which is also physiologically regulated) to maintain reduced levels of nucleosides independently of environmental factors. In our experiments, the fact that dKO mice presented higher basal levels of nucleosides than the previous generations might have influenced our results. Since the starting basal levels were higher, the reduction necessary to reach WT levels had to be more pronounced. We can speculate that, if the basal levels of dKO mice had been the same as the one observed in previous studies, maybe the reduction observed in this work would have reached WT levels.

The nucleoside basal levels also presented a noticeable individual variability. For this reason, it was important to compare the basal and post-treatment values of each animal separately. Noticeably, all animals treated with LNPs and AAV vectors presented a significantly higher basal nucleoside levels than the mean. Although the reasons are unknown, it is important to note that all of them belonged to the same litter of the dKO colony, while animals of the other experimental conditions were from other litters. Since this higher nucleoside basal level was observed in all animals of the group, independently of their gender, we believe that this phenomenon is probably related to this specific litter characteristics or conditions. It is noteworthy that this phenomenon was not only

observed in this litter, as others also presented a higher basal nucleoside levels than the mean. This supports the idea that the litter factor plays an important role in the basal nucleoside levels observed. In this regard, it should be noted that we observed that basal nucleoside levels correlated with the age of sample recollection (data not shown). All mice were 7-10-week old upon treatment administration, and the pre-treatment blood samples were extracted one week before injection. We observed that younger mice (6-week old) presented higher nucleoside basal levels than older mice (9-week old), although this difference was only significant in Thd (data not shown). Pre-treatment blood samples of mice treated with LNPs and AAV vectors were extracted when they were 6-week old, so the age could partly explain their higher basal plasma levels. However, it is clear that age is not the main determining factor, as this nucleoside plasma level decrease was not observed over time in the negative control dKO group. It is likely that this correlation is not causative. Overall, since data was analysed as paired values, this litter variability did not have any effect on the comparison of results between the different experimental conditions.

Even with the limitations mentioned above, the important point is that our results demonstrate that Thd and dUrd plasma levels are consistently reduced when animals are treated with LNPs carrying CRISPR/Cas9 and rAAV2/8 vectors carrying the templates. Due to time and budget limitations, we were only able to test one dose of this combination. In the future, it would be interesting to test other doses. As the therapeutic effect is expected to be dose-dependent, it would be worth trying to determine if higher doses are capable of further reducing the nucleoside plasma concentrations to reach WT levels. Since there are two elements to optimize, there are multiple combinations of LNP and rAAV2/8 vector doses that could be tested.

The timing of administration of both LNPs and rAAV2/8 vectors could be another point of optimization for future experiments. In the present work, mice were first injected with the rAAV2/8 vectors containing the DNA templates. The next day, LNPs carrying the CRISPR/Cas9 RNAs were administered. We choose this experimental setup based on the very transient nature of the LNPs effect. As many studies have proven [205,270,271,279,280,281,282,283], AAV vectors remain inside transduced cells for long periods of time. Within a day after an intravenous injection, AAV genomes are already present in the nucleus, despite the fact that nuclear AAV transport is a slow and inefficient process [155,162,163]. On the other hand, the peak of expression of mRNAs delivered by LNPs is at 4 hours post-administration, but this expression is transient and disappears within days [309,312]. For these reasons, we decided to first administer the templates, and later the CRISPR/Cas9 molecules. This way, we estimated that the effect of CRISPR/Cas9 would overlap with the presence of DNA templates inside the nucleus, as they have a much slower kinetics. However, we only tested this schedule, so it is possible that other timings would yield better results. A more delayed administration of CRISPR/Cas9 could also be an option, so that more template copies would be available in the nucleus. We can also speculate about the possibility of re-administering the LNPs carrying CRISPR/Cas9 until the desired therapeutic effect is obtained. The nucleoside plasma levels could be strictly monitored after each injection, and more LNPs would be administered until they reached WT levels. Since the templates remain inside the nucleus long-term, the window of time in which re-administration would be possible is relatively wide (several weeks). In theory,

this approach would be feasible, as LNPs are not expected to be immunogenic. However, Cas9 protein can be immunogenic [452,453,451], so this strategy must be carefully planned, as a sustained Cas9 expression in edited cells could trigger their immune destruction. In this regard, the fact that transgene expression is restricted to hepatocytes could facilitate the induction of immune tolerance [199,200,201,202,203,204,205,206,207].

Since LNPs exhibited good results in hepatocyte delivery *in vivo*, an approach where the DNA templates are also carried by them could be explored. One of the main limitations of AAV vectors is the immunogenicity of capsids [151,196,243]. The immune response triggered during the first AAV vector exposition causes the development of humoral and sometimes cellular immune responses to capsid antigens. For this reason, a second administration of the same AAV vector would be very ineffective. On the other hand, LNPs are not immunogenic per se, so they can be re-administered multiple times. Even if the first administration is not sufficient to achieve clinical improvement, patients would be able to be treated with a second dose to increase the therapeutic effect. It is important to note that the LNPs used in this work are optimized for mRNA delivery, so it is likely that this formulation is not the best one for carrying dsDNA molecules. Until now, Acuitas Therapeutics® has mainly focused on siRNA [306,307,308] and mRNA delivery [309,310,311,312], but maybe it could be interesting to open a new experimental research line on DNA delivery. In this regard, a major limitation of non-viral vectors in general is their inability to deliver nucleic acids into the nucleus of target cells [506]. There is no need of nuclear entry for mRNA molecules, as they are translated in the cytoplasm. But DNA needs to be located inside the nucleus to be expressed or even when used as template in a gene editing approach like ours. To overcome this issue, several studies have explored the use of cell penetrating or NLS peptides in the LNP formulations [507,508,509]. These small cationic peptides are through to interact with the anionic DNA molecules and enable their nuclear translocation. Perhaps the addition of these short peptides in the LNP formulation of Acuitas Therapeutics® could be feasible and advantageous for DNA delivery.

Although treated animals did not present any sign of acute toxicity, the presence of a mild hepatotoxicity cannot be ruled out. We did not measure plasma activity of AST and ALT enzymes, potential indicators of hepatocyte damage, but given the positive results observed, we can reasonably discard the presence of a strong T-cell immune response, which eventually would have resulted in the elimination of the edited cells.

One of the main objectives of this gene editing approach is to overcome the loss of transgene expression over time that occurs with AAV liver-directed therapies. A conventional gene therapy approach for MNGIE using rAAV2/8 vectors targeting the liver has proven to be effective in the preclinical setting [314], but a follow-up study demonstrated that at some point the transgene expression began to decrease [317]. At the end of the study, this phenomenon was observed in all treated animals, regardless of the doses administered, although animals treated with the highest AAV dose still maintained the nucleoside levels within the normal WT range. An important factor that may influence the durability of transgene expression is the promoter. In the study mentioned above, the liver-specific TBG promoter was used. In another study, the authors tested additional promoters and found that the liver-specific AAT promoter yielded better results [318]. However, they only reported that the biochemical correction was maintained after 34 weeks, so whether the loss of transgene expression also takes place with this promoter in

the longer term remains to be determined. But even then, as mentioned above, these dKO mice are not the most suitable animal models for proving long term correction, as their lifespan is not comparable to humans. The fact that mice begin to lose expression at the end of their life could translate to a loss of expression after just some years in humans. Another prime example of transgene expression loss over time is the clinical trial for haemophilia A. A 3-year follow-up study demonstrated that FVIII expression decreased over time [283]. To overcome this limitation, we propose that gene editing would be a suitable alternative to conventional AAV-based gene therapies. Due to time limitations, in this work we only present the results of a short follow-up. We hypothesize that the biochemical correction obtained with gene editing will be permanent, as other studies suggest [460,424], but with the current data we cannot conclude that this will be the case. In this regard, as stated before, a cohort of mice treated with LNPs and rAAV2/8 vectors is currently being monitored. In a longer follow-up we will be able to determine whether the levels of biochemical correction changes over time. To this end, the current animal model is appropriate, as the plasma levels of nucleosides reflect the overall TP activity very accurately. In conclusion, in the present work we cannot demonstrate that this gene editing approach attains a permanent and stable biochemical correction, but we hope that we will be able to corroborate or refute this hypothesis in the near future.

The liver structure and composition must be taken into account in liver-directed therapies like this one. When NPs or AAV vectors are administered through intravenous injection, the first hepatic cells that they encounter are Kupffer cells (liver macrophages). These cells, together with liver sinusoidal endothelial cells (LSECs), play an important role in the elimination of vectors and restrict their access to hepatocytes [510,511]. Hepatocytes have a sinusoidal (basolateral) membrane in close contact with blood circulation. This membrane has microvilli that exhibit surface receptors important for LNP recognition, like the low-density lipoprotein receptor (LDLR) and the asialoglycoprotein receptor (ASGPR) [305]. These cells are tasked with many metabolic functions. To maximize the efficiency of liver as a whole, hepatocytes distribute their work based on their position along the porto-central axis of the liver lobule. Depending on the location of hepatocytes within the liver (periportal or perivenous), they are specialized in different metabolic pathways. This differential specialization also affects the type of non-parenchymal cell subtypes within the liver microenvironment. This phenomenon is known as liver zonation. It is an important factor to take into account for liver-directed therapies. There are some diseases that affect only one type of hepatocytes, so the therapies need to target this group specifically. Moreover, this metabolic zonation can vary among species, stages of life and liver diseases [512,513,514,506]. All these factors play a role in the efficiency of vector delivery to hepatocytes. In our case, liver zonation is also important for the differential gene expression patterns observed between periportal and perivenous hepatocytes. Alb is expressed in all hepatocytes, but the major expression is found in the periportal zone [512,515]. Consequently, the gene editing strategy proposed in this work for this locus should preferentially target the periportal hepatocytes to maximize the therapeutic effect. Generally, most intravenously injected nucleic acid therapeutics predominantly target perivenous cells [506]. In mice, rAAV2/8 vectors predominantly transduce perivenous hepatocytes, while in non-human primates they mainly transduce periportal hepatocytes [516]. This fact could have impacted negatively in our results, as we used a murine model. In the future it would be worth to determine the liver zone preference that exhibits the

transfection by LNPs from Acuitas Therapeutics®. In our gene editing approach both DNA templates and CRISPR/Cas9 molecules need to be introduced into the same cells, so it is important to study their particular liver zonation. It would also be interesting to study the transfection preference of KH-ret from GEMAT-IQS. Maybe the poor gene editing results obtained with these PNPs could be partially explained by the fact that they do not transduce the same type of hepatocytes as rAAV2/8 vectors.

Overall, we find these results promising, although there is a lot of room for optimization.

6.4. The biochemical correction was similar in both genders

Both genders were represented in all the experimental conditions tested. The decision to include both genders was influenced by many factors, but the main one was the limited time available to perform all *in vivo* experiments. As this project was started from scratch, much time was spent in designing the optimal tools and testing the feasibility of the gene editing approach *in vitro*. For this reason, little time was left for the *in vivo* experimentation. Moreover, our group did not have the dKO colony available from the beginning. Dr. Ramon Martí group kindly provided us with three dKO mice to start our own colony. However, as mentioned above, the housing facilities were changed recently, and by the end of 2019 mice were still adapting to the new environmental conditions. This caused problems in their reproduction, so at first the number of animals available for experimentation was very limiting. This problem improved over time, but it was hard to start our colony. For this reason, we decided to use all animals available, regardless of their gender.

Although the use of both genders added more variability to the study, we were able to improve one of the three R principles, the reduction of animals. When using only one gender, approximately half of the newborns have to be discarded, which consequently increases the total number of animals used. In our case all newborns were used for experimentation and we maximized the information obtained per animal. This also allowed the detection of differences of efficiency based on gender. Moreover, including animals from both genders makes the sample more representative of the MNGIE patients, as this disease affects equally males and females.

It is important to note that there was a difference in basal Thd and dUrd plasma levels of dKO mice depending on the gender. Females presented higher nucleoside levels than males. Interestingly, this phenomenon was not observed in WT mice (Fig. 49). The reason for this gender difference is unknown at present. Previous studies only used males to determine the efficacy of treatments, so obviously the authors did not detect this gender particularity. However, some studies described plasma metabolite differences between genders, so there are precedents of this phenomenon [517,518].

With the results we obtained, we can conclude that gender was not a critical factor in our experimental conditions. Animals treated with LNPs, which exhibited the best results, had similar outcomes independently from their gender. The difference in basal levels between genders has to be taken into account when analysing the data obtained in our work. If we only look at the post-treatment Thd and dUrd plasma values, the lowest values correspond

to males. But since males also presented lower basal values, it would not be appropriate to conclude that the efficiency of the gene editing was lower in females. To compare both genders we analysed the reduction observed post-treatment as paired values, and in this case we concluded that the decrease tendency was equivalent in both genders. These results may be counterintuitive, as it is known that AAV vectors administration results in a 5-fold higher liver transgene expression in male mice. One study suggested that this gender difference was due to the fact that AAVs transduce hepatocytes via an androgen-dependent pathway [213], while another one proposed that this sex-specific pattern was due to post-transduction factors [208]. Our results are in accordance with this second study, because if the transduction pathway was the main determining factor, we would have observed higher efficiency in males. Following this hypothesis, this gender difference would only be observed when the AAV genomes need to be expressed. In the first gene editing approach (CRISPR/Cas9 carried by NPs), AAV vectors only carry the templates and do not need to be expressed, so the sex-specific pattern should not affect in this case. On the other hand, this phenomenon would have an effect in the second gene editing approach (CRISPR/Cas9 carried by rAAV2/8 vectors), because in this case the AAV genomes would need to be expressed. Since the results obtained with this second approach were poor, we cannot conclude the presence or absence of this sex-specific pattern in this case. On the other hand, there is no evidence that LNPs function better in a particular gender, so in this case a gender difference was not expected and our results corroborated this. Another possible explanation of not detecting differences between genders is that the limiting factor of our approach was the LNPs efficacy and not rAAV2/8 vector transduction. If that was the case, we should start by optimizing the dose of LNPs to maximize the gene editing efficiency.

6.5. The biochemical correction was effective in both loci

In this work two different loci were tested as genomic targets for gene editing. The first one was the endogenous *Tymp* locus. Although the promoter of this gene is not particularly strong, placing the cDNA under the control of their own promoter had its advantages, as their expression would assimilate physiological regulation. On the other hand, we decided to test a stronger promoter, in case the endogenous *Tymp* promoter was not able to express the cDNA with enough potency. In this regard, we opted for *Alb* locus, a gene that is strongly expressed in hepatocytes. Together with haemoglobin, albumin is amongst the most expressed proteins in humans (an average adult produces 15 g/day). In this case, a hybrid Alb-hTP protein containing the secretory peptide of Alb would be generated, so this hybrid enzyme would be mainly secreted to plasma, as confirmed by our results [312,343,344,345,346].

Although the two strategies were quite different, the biochemical correction observed in animals injected with LNPs and rAAV2/8 vectors was similar in both loci. Even if the overall level of biochemical correction was comparable, the factors that determined these results were different in both cases.

The cutting efficiency of CRISPR/Cas9 delivered by LNPs was quite different in both loci (measured by NGS). In the *Alb* locus, the cutting efficiency was approximately 50%, while

in the *Tymp* locus it was much lower (Fig. 46). It is well known that every gRNA has a different targeting efficiency, so part of the difference observed could be attributed to the particular design of the gRNAs of this work [329,406]. Another important factor is that the genomic location and expression patterns of both loci are very different. Since *Alb* gene is highly expressed in our target cells [512,515,519], it is likely that the chromatin in this locus was more exposed to the nuclease.

The design of the DNA templates was also quite different in both cases. In the case of *Tymp* locus, the template only contained a partial m*Tymp* cDNA (without exons 1-2), while the template of *Alb* locus contained a resected h*TYMP* cDNA (without the first 15 nucleotides) flanked by HAs. These intrinsic differences surely had an impact in the insertion efficiency. Unfortunately, we were unable to quantify the insertion efficiency due to the low sensitivity of the technique used, so we could not determine which loci yielded the best insertion results.

In the edited cells, mRNA expression patterns were also expected to differ between loci. The *Alb* promoter would ensure an overexpression of TP mRNA, while *Tymp* promoter would generate a more tamed expression. The TP protein generated in the edited cells would also be different. In the *Tymp* locus, a successful insertion of the template would generate WT mTP molecules. Meanwhile, in the *Alb* locus, hybrid Alb-hTP molecules would be synthesized, and a large proportion of them would probably secreted into circulation. Due to its hybrid nature, this enzyme could be less active than its WT counterpart, although it would gain the ability to be secreted and act directly on the nucleosides accumulated in plasma.

Taking all this into account, we hypothesize that the biochemical correction observed in plasma was similar in both loci because different factors balanced each other. While in the *Tymp* locus the cDNA expression was lower, the resulting TP molecules were likely fully functional. In the *Alb* locus the cDNA was overexpressed, but the hybrid Alb-hTP molecules could have lost part of their enzymatic functionality. Moreover, even if they retained activity when secreted into plasma, their turnover would probably be faster in the extracellular environment.

Altogether, it should be noted that there are many factors that differ between the gene editing strategies of both loci, so it would not be correct to infer that one locus is better than the other taking into account only the results obtained.

6.6. Hybrid Alb-hTP was functional upon secretion

The choice of the *Alb* locus as an alternative to the endogenous *Tymp* locus was inspired by other studies, some of which have already reached clinical stages [312,343,344,345,346]. Since Alb exon 1 encodes a secreted peptide that is cleaved from the final product, the addition of the cDNA with a splice acceptor site into intron 1 allows the creation of a new protein combining the secretory peptide and the protein of interest. The correct editing in this locus would preclude normal Alb expression, but we assumed that the efficiency would not be high enough to significantly reduce plasma levels of this protein, or at least that such a reduction would not be clinically relevant. Since the production of Alb is not

essential for hepatocyte viability, its absence would not affect cellular functions or represent a selective disadvantage.

It is important to note that, up to now, this strategy has only been tested with proteins that are naturally secreted (like FIX in haemophilia B), while TP is an intracellular enzyme. Initially, we didn't know whether a secreted form of TP would be functional or very short-lived, but we decided to take a chance. One study suggested that this enzyme was secreted by tumour cells [520], and another one reported the presence of TP in plasma of cancer patients [521]. Another group studied the pharmacokinetics and tissue distribution of TP after intravenous injection, and concluded that this enzyme was stable and presented a relatively long half-life in the circulation [522]. All these studies led us to believe that a secreted TP enzyme could still be active in plasma. If that was the case, the secreted enzyme would be able to directly help in the clearance of the systemic nucleoside accumulation. Moreover, its presence in circulation would facilitate their access to other tissues. In this regard, the pathophysiology of MNGIE is very relevant, as it involves a systemic accumulation of hydrosoluble and highly diffusible molecules (nucleosides) that results in mitochondrial toxicity. However, experimental evidences suggest that the defect can probably be corrected at basically any level [53,77,88,313,314,315,316,317,318].

In physiological conditions, the endogenous TP is an intracellular homodimeric enzyme. The newly translated TP contains a pro-peptide that is eliminated during maturation. Then, the mature peptides combine between each other and form the active enzymatic dimeric molecules. On the other hand, Alb is an extracellular protein. The newly translated Alb contains a secretory peptide that only acts as a signal and is excised from the final Alb protein when secreted. With this information, it is likely that the secreted hybrid enzyme also lacks this secretory peptide, so our secreted TP form likely has a very similar structure as that of the WT TP.

Since Alb-hTP is a new hybrid created artificially, there was no previous information of it, so we had to empirically test its functionality. For this reason, we generated the pCMV-Alb-TP plasmid, which contained the hybrid gene under the control of the CMV promoter. With this plasmid, we demonstrated that the hybrid was stable *in vitro*. Not only was the protein detectable by Western blot, but we also detected some enzymatic activity. These results suggested that the hybrid protein was successfully transcribed and translated. This hybrid retained some intracellular activity, which means that these molecules were able to dimerize even if they were not identical to WT TP. However, it is important to note that the intracellular activity detected was quite low. The presence of the Alb secretory peptide probably reduced the normal functionality of TP. But the ability of this hybrid to be secreted also contributed to the low intracellular activity. The fact that we detected enzymatic activity in culture medium also supports this notion. It was not possible to formally demonstrate if the secreted protein had a crippled functionality or not. The WT TP is not normally found in medium, so there was no point of comparison. Altogether, the *in vitro* results confirmed the viability of the *Alb* gene editing approach.

The *in vivo* results obtained further support the feasibility of the approach. The biochemical correction observed in mice treated with LNPs and rAAV2/8 vectors is a proof that the edited cells were able to generate functional TP. The intracellular hepatic activity detected was very low, but these animals presented a robust TP activity in plasma.

As previously said, since Alb-hTP can be secreted, the low intracellular activity can be explained by the fact that most molecules are rapidly released into circulation. The high activity values detected in plasma corroborate this hypothesis. Since dKO mice present undetectable TP activity in plasma (and in blood), we can conclude that the plasma activity TP detected in those edited mice was due to the Alb-hTP molecules secreted.

We think that in the future it can be worth to further investigate the impact of the Alb secretory peptide in TP structure and functionality. The low intracellular activity observed can be partially explained by the Alb-hTP secretory nature, but it remains to be determined if the structure of the hybrid also plays some role in the decreased activity. The addition of the secretory peptide from Alb in their structure may have had an impact on the dimer formation. Moreover, it would be interesting to determine if the hybrid dimerizes before or after being secreted. The detection of some intracellular activity suggests that at least some dimerization occurs before secretion, but further studies are needed. The stability and lifespan of the hybrid in circulation must also be studied. Although a study concluded that TP was stable and presented a relatively long half-life in the circulation [522], we cannot assure that it is the same with the hybrid. Maybe an improvement in the enzyme stability could yield higher plasma activity levels. Perhaps with the help of protein engineering the enzyme stability and functionality could be further optimized.

In contrast to the gene editing approach targeting the *Tymp* locus, Alb-hTP expression from the *Alb* locus is restricted to the edited hepatocytes. If we assume that most of the synthesized enzyme is secreted into circulation and that it does not significantly extravasate, as the distribution volume (plasma volume) is known, the reduction in nucleoside levels will be basically dependent on plasma enzyme activity, two parameters that can be measured. With these premises, we hypothesize that it will possible to create mathematical models that will allow to make precise predictions on the doses of rAAV2/8 vectors and LNPs that will be required to achieve a specific level of correction on a per case basis.

Altogether, although we can conclude that Alb-hTP hybrid retains activity in plasma, more studies are needed to determine if this form is suitable for future clinical applications. If this was demonstrated, the use of *Alb* locus for gene editing could become a platform to treat many monogenic diseases. This promoter could be used to synthesize and secrete a wide range of proteins. In this regard, this could not only be useful for proteins that are naturally secreted, but also with intracellular proteins that can function in the extracellular medium, which is the case of TP. Diseases that are now treated with enzyme replacement therapies could be suitable targets for this platform. However, since hybrid proteins are generated with this approach, it would be necessary to analyse each hybrid to ascertain that it retains its natural function.

6.7. Gene editing can occur without CRISPR/Cas9

Unexpectedly and very surprisingly, our results indicate that the presence of templates alone was sometimes enough to trigger an efficient gene editing. In other words, under some conditions, the presence of CRISPR/Cas9 was not essential for an efficient gene

editing. The template insertion without a nuclease per se was not totally unexpected, what was surprising was the high percentage of cells that were edited. Since templates contained some homology with the target loci, they could be inserted successfully by HR. However, the presence of nucleases like CRISPR/Cas9 was expected to markedly increase the chances of insertion [95,319,320,322]. This phenomenon was mainly observed in the *Alb* locus, but this could be due to the fact that TP mRNA and the efficiency of gene insertion could only be assessed in this locus. Consequently, we have more data of the *Alb* gene editing results. We cannot rule out the possibility that gene editing in the *Tymp* locus also occurs with only the presence of DNA templates.

During *in vitro* experimentation, we found that cells transfected with plasmids containing the templates (without nuclease) displayed some level of gene editing. The PCRs with specific primers yielded bands of the expected size, and the expected insertion of these templates was confirmed by Sanger sequencing. When these bands were quantified, the percentages of template insertion obtained were similar to those obtained in cells also transfected with CRISPR/Cas9 RNAs. This observation was so unexpected that the experiment was repeated (and the results confirmed) several times. One could argue that CRISPR/Cas9 may not have worked in these experiments, for whatever reasons. But this is highly unlikely, as the same results were obtained several times, and we checked that CRISPR/Cas9 transfection caused DSBs in the correct target sequences periodically. Moreover, this does not change the fact that DNA templates alone could trigger gene editing. At this point, it is important to note that this method of insertion detection has many limitations. We could easily detect the band that corresponded to HDR insertion by PCRs, as its size remained mostly stable, but the insertion through NHEJ was trickier, as it could yield bands of different size. This method of quantification is not very accurate, it only gives an approximation. We also detected some hTP mRNA expression in those cells, which corroborate the fact that cells were correctly edited and it was not an artefact. One possible explanation to this is that Hepa1c1c7 cells have a highly active HDR-repair pathway, although this has not been reported to date.

In addition, the *in vivo* results further confirmed this phenomenon. Although animals treated with AAV templates alone did not have any detectable plasma nucleoside decrease, we detected insertion of the templates in some liver samples, and they even showed detectable TP mRNA expression. These mRNA results are especially shocking, as these mice presented higher TP mRNA expression than animals treated additionally with the KH-ret PNP formulation carrying the CRISPR/Cas9 RNAs. As discussed above, the KH-ret formulation was toxic to mice, so this could partially explain the fact that these mice had a worse outcome. We could speculate that these PNPs caused hepatic toxicity relatively rapidly upon administration, perhaps destroying cells that were already transduced by rAAV2/8 vectors carrying the templates. On the other hand, animals treated with LNPs and rAAV2/8 templates did show better outcomes than those treated only with the templates. At least in this experimental condition, the presence of CRISPR/Cas9 was determinant in increasing the gene editing efficiency.

Interestingly, the insertion of template without nucleases was reported in one gene editing study also targeting the *Alb* locus [523]. In this case, they used AAV vectors carrying a promoter-less DNA template containing the FIX cDNA flanked by HAs homologous to the *Alb* locus. Although this approach is similar to the one we used, in this case the cDNA was

intended to be introduced after the stop codon of *Alb*. The template also included a 2A-peptide that allowed the full translation of both Alb and FIX as separate proteins. In this study no nuclease was used, but gene editing was still detected. Although the efficiency results they obtained were not spectacular, the percentage of edited cells was enough to partially correct the bleeding phenotype of haemophilia B model mice. The authors argue that the high transcriptional activity at the *Alb* locus might make it more susceptible to transgene integration by HDR. However, these results contrast with the ones obtained in other studies that also used the *Alb* promoter, in which the presence of a nuclease was determinant [312,343,344,345,346]. Perhaps the particular structure of the DNA templates used in each work can explain the differences in the results [524]. More recently another group obtained similar results using this same design for their DNA templates [525]. In this case, the study was performed in a mouse model of Crigler-Najjar syndrome type I (CNSI). Neonatal transduction with the templates resulted in the complete rescue from neonatal lethality and achieved a therapeutic reduction in plasma bilirubin at long-term. The results published in these two studies are in accordance with our results. It is possible that the templates we designed were also able to be introduced by HR in a relatively efficient fashion without the presence of DSBs. Indeed, gene editing without nuclease has potential advantages. The expression of Cas9 can induce the immune clearance of the edited cells [451,452,453], and the possible off-target effects can cause mutagenesis [388,388,389,390]. Although the use of CRISPR/Cas9 in RNA format probably reduces those risks, a therapy that completely avoids its use may be even better in terms of safety.

Overall, the insertion of templates in the absence of CRISPR/Cas9 was an unexpected discovery that poses its advantages. In the future it would be interesting to further investigate this phenomenon.

7. CONCLUSIONS

1. Gene editing approaches targeting the *Tymp* and the *Alb* loci using CRISPR/Cas9 are feasible both *in vitro* and *in vivo*.

2. The DNA templates designed were successfully integrated in both loci by either HDR or NHEJ.

3. The hybrid Alb-hTP protein generated by gene editing in the *Alb* locus was functional and maintained its enzymatic activity upon secretion.

4. In our experimental conditions, the LNP formulation was more efficient in transfecting hepatocytes than the KH-ret formulation.

5. Under our experimental conditions, delivering the CRISPR/Cas9 system by LNPs in RNA form was more efficient than in its DNA form delivered by rAAV2/8 vectors.

6. The gene editing approach using LNPs to deliver the CRISPR/Cas9 system and rAAV2/8 vectors to deliver the DNA templates was the only strategy that attained a robust and consistent reduction of Thd and dUrd plasma levels in the murine model of MNGIE.

7. The livers of mice treated with LNPs and rAAV2/8 vectors produced the highest levels of TP mRNA and functional TP protein.

8. Mice treated with LNPs and rAAV2/8 vectors targeting the *Alb* locus presented high levels of TP activity in plasma.

9. Under our experimental conditions, and although the two strategies were not comparable, the overall biochemical correction of mice treated with LNPs and rAAV2/8 vectors was similar in both loci.

10. Under some experimental conditions, efficient gene editing can be obtained both *in vitro* and *in vivo* in the absence of nuclease, only with the DNA template containing homology regions.

11. This study corroborates that liver gene editing was feasible both in the endogenous *Tymp* and the *Alb* loci, although there is a large room for optimization. Such approaches could not only constitute valid therapeutic options for MNGIE, but also for other monogenic diseases.

8. REFERENCES

1. Orphanet: Mitochondrial neurogastrointestinal encephalomyopathy. https://www.orpha.net/consor/cgi-bin/OC_Exp.php?lng=EN&Expert=298.

2. Pacitti, D., Levene, M., Garone, C., Nirmalananthan, N. & Bax, B. E. Mitochondrial Neurogastrointestinal Encephalomyopathy: Into the fourth decade, what we have learned so far. *Front. Genet.* **9**, 1–22 (2018).

3. Hirano, M. *et al.* Mitochondrial neurogastrointestinal encephalomyopathy (MNGIE): Clinical, biochemical, and genetic features of an autosomal recessive mitochondrial disorder. *Neurology* **44**, 721–727 (1994).

4. Garone, C., Tadesse, S. & Hirano, M. Clinical and genetic spectrum of mitochondrial neurogastrointestinal encephalomyopathy. *Brain* **134**, 3326–3332 (2011).

5. Hirano, M., Garone, C. & Quinzii, C. M. CoQ 10 deficiencies and MNGIE: Two treatable mitochondrial disorders. *Biochim. Biophys. Acta.* **1820**, 625–631 (2012).

6. Okamura, K., Santa, T., Nagae, K. & Omae, T. Congenital oculoskeletal myopathy with abnormal muscle and liver mitochondria. *J. Neurol. Sci.* **27**, 79–91 (1976).

7. Bardosi, A. *et al.* Myo-, neuro-, gastrointestinal encephalopathy (MNGIE syndrome) due to partial deficiency of cytochrome-c-oxidase. *Acta Neuropathol.* 248–258 (1987).

8. Nishino, I., Spinazzola, A. & Hirano, M. Thymidine Phosphorylase gene mutations in MNGIE, a human mitochondrial disorder. *Science.* **283**, 689–692 (1999).

9. Hirano, M. *et al.* Mitochondrial neurogastrointestinal encephalomyopathy syndrome maps to chromosome 22q13.32-qter. *Am. J. Hum. Genet.* **63**, 526–533 (1998).

10. Nishino, I. *et al.* Mitochondrial neurogastrointestinal encephalomyopathy: An autosomal recessive disorder due to thymidine phosphorylase mutations. *Ann. Neurol.* **47**, 792–800 (2000).

11. Gamez, J. *et al.* A novel thymidine phosphorylase mutation in a Spanish MNGIE patient. *J. Neurol. Sci.* **228**, 35–39 (2005).

12. Hirano, M., Nishigaki, Y. & Martí, R. Mitochondrial Neurogastrointestinal Encephalomyopathy (MNGIE): A disease of two genomes. *Neurologist* **10**, 8–17 (2004).

13. Song, S., Wheeler, L. J. & Mathews, C. K. Deoxyribonucleotide pool imbalance stimulates deletions in HeLa cell mitochondrial DNA. *J. Biol. Chem.* **278**, 43893–43896 (2003).

14. Suh, B. C. *et al.* Compound heterozygous mutations of TYMP as underlying causes of mitochondrial neurogastrointestinal encephalomyopathy (MNGIE). *Mol. Med. Rep.* **8**, 17–22 (2013).

15. Falcão de Campos, C., Oliveira Santos, M., Roque, R., Conceição, I. & de Carvalho, M. Mitochondrial Neurogastrointestinal Encephalomyopathy: Novel pathogenic mutation in Thymidine Phosphorylase gene in a patient from Cape Verde Islands. *Case Rep. Neurol. Med.* **2019**, 1–4 (2019).

16. Kocaefe, Y. C., Erdem, S., Ozguc, M. & Tan, E. Erratum: Four novel thymidine phosphorylase gene mutations in mitochondrial neurogastrointestinal

encephalomyopathy syndrome (MNGIE) patients (European Journal of Human Genetics (2003) vol. 11 (50-56)). *Eur. J. Hum. Genet.* **11**, 551 (2003).

17. Szigeti, K. *et al.* MNGIE with lack of skeletal muscle involvement and a novel TP splice site mutation. *J. Med. Genet.* **41**, 125–129 (2004).

18. Blazquez, A. *et al.* Increased muscle nucleoside levels associated with a novel frameshift mutation in the thymidine phosphorylase gene in a Spanish patient with MNGIE. *Neuromuscul. Disord.* **15**, 775–778 (2005).

19. Martí, R. *et al.* Definitive diagnosis of Mitochondrial Neurogastrointestinal Encephalomyopathy by biochemical assays. *Clin. Chem.* **50**, 120–124 (2004).

20. Ishikawa, F. *et al.* Identification of angiogenic activity and the cloning and expression of platelet-derived endothelial cell growth factor. *Nature* **338**, 557–562 (1989).

21. Hagiwara, K. *et al.* Organization and chromosomal localization of the human platelet-derived endothelial cell growth factor gene. *Mol. Cell. Biol.* **11**, 2125–2132 (1991).

22. Miyazono, K., Okabe, T., Urabe, A., Takaku, F. & Heldin, C. H. Purification and properties of an endothelial cell growth factor from human platelets. *J. Biol. Chem.* **262**, 4098–4103 (1987).

23. Friedkin, M. & Roberts, D. The enzymatic synthesis od nucleosides. Thymidine phosphorilase in mammalian tissue. *J. Biol. Chem.* 245–256 (1954).

24. Desgranges, C., Razaka, G., Rabaud, M. & Bricaud, H. Catabolism of thymidine in human blood platelets purification and properties of thymidine phosphorylase. *Biochim. Biophys. Acta.* **654**, 211–218 (1981).

25. Fox, S. B. *et al.* Platelet-derived endothelial cell growth factor/thymidine phosphorilase expression in normal tissues: an immunohistochemical study. *J. Pathol.* **176**, 183–190 (1995).

26. Martí, R., Nishigaki, Y. & Hirano, M. Elevated plasma deoxyuridine in patients with thymidine phosphorylase deficiency. *Biochem. Biophys. Res. Commun.* **303**, 14–18 (2003).

27. Matsuura, T. *et al.* Thymidine phosphorylase expression is associated with both increase of intratumoral microvessels and decrease of apoptosis in human colorectal carcinomas. *Cancer Res.* **59**, 5037–5040 (1999).

28. Asai, K. *et al.* Neurotrophic action of gliostatin on cortical neurons. Identity of gliostatin and platelet-derived endothelial cell growth factor. *J. Biol. Chem.* **267**, 20311–20316 (1992).

29. Spinazzola, A. *et al.* Altered thymidine metabolism due to defects of thymidine phosphorylase. *J. Biol. Chem.* **277**, 4128–4133 (2002).

30. Hirano, M., Martí, R., Spinazzola, A., Nishino, I. & Nishigaki, Y. Thymidine phosphorylase deficiency causes MNGIE: An autosomal recessive mitochondrial disorder. *Nucleosides, Nucleotides and Nucleic Acids* **23**, 1217–1225 (2004).

31. Valentino, M. L. *et al.* Thymidine and deoxyuridine accumulate in tissues of patients with mitochondrial neurogastrointestinal encephalomyopathy (MNGIE). *FEBS Lett.*

581, 3410–3414 (2007).

32. Lara, M. C., Valentino, M. L., Torres-Torronteras, J., Hirano, M. & Martí, R. Mitochondrial neurogastrointestinal encephalomyopathy (MNGIE): Biochemical features and therapeutic approaches. *Biosci. Rep.* **27**, 151–163 (2007).

33. Rötig, A. & Poulton, J. Genetic causes of mitochondrial DNA depletion in humans. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1792**, 1103–1108 (2009).

34. Pontarin, G., Gallinaro, L., Ferraro, P., Reichard, P. & Bianchi, V. Origins of mitochondrial thymidine triphosphate: Dynamic relations to cytosolic pools. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 12159–12164 (2003).

35. Pontarin, G. *et al.* p53R2-dependent ribonucleotide reduction provides deoxyribonucleotides in quiescent human fibroblasts in the absence of induced DNA damage. *J. Biol. Chem.* **282**, 16820–16828 (2007).

36. Van Rompay, A. R., Johansson, M. & Karlsson, A. Phosphorylation of nucleosides and nucleoside analogs by mammalian nucleoside monophosphate kinases. *Pharmacol. Ther.* **87**, 189–198 (2000).

37. Wang, L., Saada, A. & Eriksson, S. Kinetic properties of mutant human thymidine kinase 2 suggest a mechanism for mitochondrial DNA depletion myopathy. *J. Biol. Chem.* **278**, 6963–6968 (2002).

38. Richards, R. G., Sowers, L. C., Laszlo, J. & Sedwick, W. D. The occurrence and consequences of deoxyuridine in DNA. *Adv. Enzyme Regul.* **22**, 157–185 (1984).

39. Ferraro, P. *et al.* Mitochondrial deoxynucleotide pools in quiescent fibroblasts: A possible model for mitochondrial neurogastrointestinal encephalomyopathy (MNGIE). *J. Biol. Chem.* **280**, 24472–24480 (2005).

40. Pontarin, G. *et al.* Mitochondrial DNA depletion and thymidine phosphate pool dynamics in a cellular model of mitochondrial neurogastrointestinal encephalomyopathy. *J. Biol. Chem.* **281**, 22720–22728 (2006).

41. Giordano, C. *et al.* Gastrointestinal dysmotility in mitochondrial neurogastrointestinal encephalomyopathy is caused by mitochondrial DNA depletion. *Am. J. Pathol.* **173**, 1120–1128 (2008).

42. López, L. C. *et al.* Unbalanced deoxynucleotide pools cause mitochondrial DNA instability in thymidine phosphorylase-deficient mice. *Hum. Mol. Genet.* **18**, 714–722 (2009).

43. González-Vioque, E., Torres-Torronteras, J., Andreu, A. L. & Martí, R. Limited dCTP availability accounts for mitochondrial DNA depletion in mitochondrial neurogastrointestinal encephalomyopathy (MNGIE). *PLoS Genet.* **7**, (2011).

44. Garcia-Diaz, B. *et al.* Deoxynucleoside stress exacerbates the phenotype of a mouse model of mitochondrial neurogastrointestinal encephalopathy. *Brain* **137**, 1337–1349 (2014).

45. Nishigaki, Y., Martí, R. & Hirano, M. ND5 is a hot-spot for multiple atypical mitochondrial DNA deletions in mitochondrial neurogastrointestinal encephalomyopathy. *Hum. Mol. Genet.* **13**, 91–101 (2004).

46. Melnyk, S. *et al.* Uracil misincorporation, DNA strand breaks, and gene amplification

are associated with tumorigenic cell transformation in folate deficient/repleted Chinese hamster ovary cells. *Cancer Lett.* **146**, 35–44 (1999).

47. Bjursell, G. & Reichard, P. Effects of thymidine on deoxyribonucleoside triphosphate pools and deoxyribonucleic acid synthesis in Chinese hamster ovary cells. *J. Biol. Chem.* **248**, 3904–3909 (1973).

48. Alexeyev, M., Shokolenko, I., Wilson, G. & LeDoux, S. The maintenance of mitochondrial DNA integrity - Critical analysis and update. *Cold Spring Harb. Perspect. Biol.* **5**, (2013).

49. Berk, A. J. & Clayton, D. A genetically distinct thymidine kinase in mammalian mitochondria. *J. Biol. Chem.* **248**, 2722–2729 (1973).

50. Bestwick, R. K., Moffett, G. L. & Mathews, C. K. Selective expansion of mitochondrial nucleoside triphosphate pools in antimetabolite-treated HeLa cells. *J. Biol. Chem.* **257**, 9300–9304 (1982).

51. Marti, R. *et al.* Late-onset MNGIE due to partial loss of thymidine phosphorylase activity. *Ann. Neurol.* **58**, 649–652 (2005).

52. Bedlack, R. S. *et al.* MNGIE neuropathy: Five cases mimicking chrohic inflammatory demyelinating polyneuropathy. *Muscle and Nerve* **29**, 364–368 (2004).

53. Filosto, M. *et al.* Mitochondrial neurogastrointestinal encephalomyopathy. *J. Clin. Med.* **7**, 1–13 (2018).

54. Corazza, G. *et al.* MyoNeuroGastroIntestinal Encephalopathy: Natural history and means for early diagnosis. *Gastroenterology* **156**, 1525-1527.e4 (2019).

55. Teitelbaum, J. E. *et al.* Diagnosis and management of MNGIE syndrome in children: Case report and review of the literature. *J. Pediatr. Gastroenterol. Nutr.* **35**, 377–383 (2002).

56. Spagnoli, C. *et al.* Peripheral neuropathy and gastroenterologic disorders: An overview on an underrecognized association. *Acta Biomed.* **89**, 22–32 (2018).

57. Patel, R. *et al.* Mitochondrial neurogastrointestinal encephalopathy: A clinicopathological mimic of Crohn's disease. *BMC Gastroenterol.* **19**, 1–5 (2019).

58. Carlson, M. D. Recent advances in newborn screening for neurometabolic disorders. *Curr. Opin. Neurol.* **17**, 133–138 (2004).

59. Hirano, M. Mitochondrial neurogastrointestinal encephalopathy disease. *Gene Rev.* (2016) doi:10.1016/j.arcped.2014.08.006.

60. Yavuz, H. *et al.* Treatment of mitochondrial neurogastrointestinal encephalomyopathy with dialysis. *Arch. Neurol.* **64**, 435–438 (2007).

61. Wang, J. *et al.* Nutrition therapy for Mitochondrial Neurogastrointestinal Encephalopathy with homozygous mutation of the TYMP gene. *Clin. Nutr. Res.* **4**, 132 (2015).

62. Cabrera-Pérez, R. *et al.* Prospective therapeutic approaches in mitochondrial neurogastrointestinal encephalomyopathy (MNGIE). *Expert Opin. Orphan Drugs* **3**, 1167–1182 (2015).

63. Röeben, B. *et al.* Hemodialysis in MNGIE transiently reduces serum and urine levels

of thymidine and deoxyuridine, but not CSF levels and neurological function. *Orphanet J. Rare Dis.* **12**, 10–13 (2017).

64. La Marca, G. *et al.* Pre- and post-dialysis quantitative dosage of thymidine in urine and plasma of a MNGIE patient by using HPLC-ESI-MS/MS. *J. Mass Spectrom.* **41**, 586–592 (2006).

65. Ariaudo, C. *et al.* Mitochondrial neurogastrointestinal encephalomyopathy treated with peritoneal dialysis and bone marrow transplantation. *J. Nephrol.* **28**, 125–127 (2014).

66. Lara, M. C. *et al.* Infusion of platelets transiently reduces nucleoside. *Neurology* **67**, 1–3 (2006).

67. Hussein, E. Non-myeloablative bone marrow transplant and platelet infusion can transiently improve the clinical outcome of mitochondrial neurogastrointestinal encephalopathy: A case report. *Transfus. Apher. Sci.* **49**, 208–211 (2013).

68. Hirano, M. *et al.* Allogeneic stem cell transplantation corrects biochemical derangements in MNGIE. *Neurology* **67**, 1458–1460 (2006).

69. Filosto, M. *et al.* Course and management of allogeneic stem cell transplantation in patients with mitochondrial neurogastrointestinal encephalomyopathy. *J. Neurol.* **259**, 2699–2706 (2012).

70. Halter, J. P. *et al.* Allogeneic haematopoietic stem cell transplantation for mitochondrial neurogastrointestinal encephalomyopathy. *Brain* **138**, 2847–2858 (2015).

71. Peedikayil, M. C., Kagevi, E. I., Abufarhaneh, E., Alsayed, M. D. & Alzahrani, H. A. Mitochondrial neurogastrointestinal encephalomyopathy treated with stem cell transplantation: A case report and review of literature. *Hematol. Oncol. Stem Cell Ther.* **8**, 85–90 (2015).

72. Baker, M. K., Schutte, C. M., Ranchhod, N., Brittain, D. & Van Rensburg, J. E. Transient clinical improvement of a mitochondrial neurogastrointestinal encephalomyopathy-like syndrome after allogeneic haematopoietic stem cell transplantation. *BMJ Case Rep.* (2017) doi:10.1136/bcr-2016-218276.

73. Halter, J. *et al.* Allogeneic hematopoietic SCT as treatment option for patients with mitochondrial neurogastrointestinal encephalomyopathy (MNGIE): A consensus conference proposal for a standardized approach. *Bone Marrow Transplant.* **46**, 330–337 (2011).

74. Boschetti, E. *et al.* Liver as a source for thymidine phosphorylase replacement in mitochondrial neurogastrointestinal encephalomyopathy. *PLoS One* **9**, 10–15 (2014).

75. De Giorgio, R. *et al.* Liver transplantation for mitochondrial neurogastrointestinal encephalomyopathy. *Ann. Neurol.* **80**, 448–455 (2016).

76. D'Angelo, R. *et al.* Liver transplant reverses biochemical imbalance in mitochondrial neurogastrointestinal encephalomyopathy. *Mitochondrion* **34**, 101–102 (2017).

77. Yadak, R., Smitt, P. S., van Gisbergen, M. W., van Til, N. P. & de Coo, I. F. M. Mitochondrial neurogastrointestinal encephalomyopathy caused by thymidine phosphorylase enzyme deficiency: From pathogenesis to emerging therapeutic

options. *Front. Cell. Neurosci.* **11**, (2017).

78. De Vocht, C. *et al.* Assessment of stability, toxicity and immunogenicity of new polymeric nanoreactors for use in enzyme replacement therapy of MNGIE. *J. Control. Release* **137**, 246–254 (2009).

79. Levene, M. *et al.* Preclinical toxicity evaluation of erythrocyte-encapsulated thymidine phosphorylase in BALB/c mice and beagle dogs: An enzyme-replacement therapy for mitochondrial neurogastrointestinal encephalomyopathy. *Toxicol. Sci.* **131**, 311–324 (2012).

80. Moran, N. F., Bain, M. D., Muqit, M. M. K. & Bax, B. E. Carrier erythrocyte entrapped thymidine phosphorylase therapy for mngie. *Neurology* **71**, 686–688 (2008).

81. Levene, M. *et al.* Safety and efficacy of erythrocyte encapsulated thymidine phosphorylase in Mitochondrial Neurogastrointestinal Encephalomyopathy. *J. Clin. Med.* **8**, 457 (2019).

82. Bax, B. E., Bain, M. D., Filosto, M., Tonin, P. & Moran, N. Clinical and biochemical improvements in a patient with MNGIE following enzyme replacement. *Neurology* **81**, 1269–1271 (2013).

83. Bax, B. *et al.* Erythrocyte encapsulated Thymidine Phosphorylase for the treatment of patients with Mitochondrial Neurogastrointestinal Encephalomyopathy: Study protocol for a multi-centre, multiple dose, open label trial. *J. Clin. Med.* **8**, 1096 (2019).

84. Cámara, Y. *et al.* Administration of deoxyribonucleosides or inhibition of their catabolism as a pharmacological approach for mitochondrial DNA depletion syndrome. *Hum. Mol. Genet.* **23**, 2459–2467 (2014).

85. el Kouni, M. H., Naguib, F. N. M. & Naguib, F. N. M. Differences in activities and substrate specificity of human and murine pyrimidine nucleoside phosphorylases: Implications for chemotherapy with 5-Fluoropyrimidines. *Cancer Res.* **53**, 3687–3693 (1993).

86. Haraguchi, M., Tsujimoto, H., Fukushima, M. & Higuchi, I. Targeted deletion of both Thymidine Phosphorylase and Uridine Phosphorylase and consequent disorders in mice. *Mol. Cell. Biol.* **22**, 5212–5221 (2002).

87. Cao, D., Leffert, J. J., McCabe, J., Kim, B. & Pizzorno, G. Abnormalities in uridine homeostatic regulation and pyrimidine nucleotide metabolism as a consequence of the deletion of the uridine phosphorylase gene. *J. Biol. Chem.* **280**, 21169–21175 (2005).

88. Yadak, R. *et al.* Transplantation, gene therapy and intestinal pathology in MNGIE patients and mice. *BMC Gastroenterol.* **18**, 1–6 (2018).

89. Rylova, S. N., Mirzaee, S., Albertioni, F. & Eriksson, S. Expression of deoxynucleoside kinases and 5′-nucleotidases in mouse tissues: Implications for mitochondrial toxicity. *Biochem. Pharmacol.* **74**, 169–175 (2007).

90. Thorne, B., Takeya, R., Vitelli, F. & Swanson, X. Gene Therapy. *Adv. Biochem. Eng. Biotechnol.* (2016) doi:10.1007/10.

91. Raper, S. E. *et al.* Fatal systemic inflammatory response syndrome in a ornithine transcarbamylase deficient patient following adenoviral gene transfer. *Mol. Genet.*

Metab. **80**, 148–158 (2003).

92. McCormack, M. P. & Rabbitts, T. H. Activation of the T-Cell oncogene LMO2 after gene therapy for X-linked Severe Combined Immunodeficiency. *New English J. Med.* **350**, 913–922 (2004).

93. Rosenberg, S. A., Aebersold, P. & Cornetta, K. Gene transfer into humans - Immunotherapy of patients with advanced melanoma, using tumor-infiltrating lymphocytes modified by retroviral gene transduction. *New English J. Med.* **323**, 570–578 (1990).

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

95. Gruntman, A. M. & Flotte, T. R. The rapidly evolving state of gene therapy. *FASEB J.* **32**, 1733–1740 (2018).

96. Gene Therapy Clinical Trials Worldwide. http://www.abedia.com/wiley/vectors.php.

97. Shahryari, A. *et al.* Development and clinical translation of approved gene therapy products for genetic disorders. *Front. Genet.* **10**, 1–25 (2019).

98. Gene Therapy Clinical Trials Worldwide. http://www.abedia.com/wiley/indications.php.

99. Kawabata, K., Takakura, Y. & Hashida, M. The fate of plasmid DNA after intravenous injection in mice: Involvement of Scavenger receptors in its hepatic uptake. *Pharmaceutical Research*. **12**, 825–830 (1995).

100. Heilbronn, R. & Weger, S. Viral vectors for gene transfer: Current status of gene therapeutics. *Handb. Exp. Pharmacol.* **197**, 55–86 (2010).

101. Gómez-Aguado, I. *et al.* Nanomedicines to deliver mRNA: State of the art and future perspectives. *Nanomaterials* **10**, 1–42 (2020).

102. Telesnitsky, A. Retroviruses: Molecular biology, genomics and pathogenesis. *Futur. Virol* **5**, 539–543 (2010).

103. Poletti, V. & Mavilio, F. Interactions between retroviruses and the host cell genome. *Mol. Ther. - Methods Clin. Dev.* **8**, 31–41 (2017).

104. Parker, R. C., Varmus, H. E. & Bishop, J. M. Cellular homologue (c-src) of the transforming gene of Rous sarcoma virus: Isolation, mapping, and transcriptional analysis of c-src and flanking regions. *Proc. Natl. Acad. Sci. U. S. A.* **78**, 5842–5846 (1981).

105. Swanstrom, R., Parker, R. C., Varmus, H. E. & Bishop, J. M. Transduction of a cellular oncogene: The genesis of Rous sarcoma virus. *Proc. Natl. Acad. Sci. U. S. A.* **80**, 2519–2523 (1983).

106. Miller, A. D., Jollyt, D. J., Friedmannt, T. & Verma, I. M. A transmissible retrovirus expressing human hypoxanthine phosphoribosyltransferase (HPRT): Gene transfer into cells obtained from humans deficient in HPRT. *Biochemistry* **80**, 4709–4713 (1983).

107. Stuhlmann, H., Jaenisch, R. & Mulligan, R. C. Construction and properties of

replication-competent murine retroviral vectors encoding methotrexate resistance. *Mol. Cell. Biol.* **9**, 100–108 (1989).

108. Elsner, C. & Bohne, J. The retroviral vector family: something for everyone. *Virus Genes* **53**, 714–722 (2017).

109. Gordon, E. M. & Hall, F. L. Rexin-G, a targeted genetic medicine for cancer. *Expert Opin. Biol. Ther.* **10**, 819–832 (2010).

110. Aiuti, A., Roncarolo, M. G. & Naldini, L. Gene therapy for ADA-SCID, the first marketing approval of an ex vivo gene therapy in Europe: paving the road for the next generation of advanced therapy medicinal products. *EMBO Mol. Med.* **9**, 737–740 (2017).

111. Ciceri, F. *et al.* Infusion of suicide-gene-engineered donor lymphocytes after family haploidentical haemopoietic stem-cell transplantation for leukaemia (the TK007 trial): a non-randomised phase I-II study. *Lancet Oncol.* **10**, 489–500 (2009).

112. Roberts, Z. J., Better, M., Bot, A., Roberts, M. R. & Ribas, A. Axicabtagene ciloleucel, a first-in-class CAR T cell therapy for aggressive NHL. *Leuk. Lymphoma* **59**, 1785–1796 (2017).

113. Maude, S. L. *et al.* Tisagenlecleucel in children and young adults with B-cell lymphoblastic leukemia. *N. Engl. J. Med.* **378**, 439–448 (2018).

114. Lee, B. Invossa, a first-in-class of cell and gene therapy for osteoarthritis treatment: the phase III trial. *Osteoarthr. Cartil.* **26**, S43–S44 (2018).

115. Miller, D. G., Adam, M. A. & Miller, A. D. Gene transfer by retrovirus vectors occurs only in cells that are actively replicating at the time of infection. *Mol. Cell. Biol.* **10**, 4239–4242 (1990).

116. Durand, S. & Cimarelli, A. The inside out of lentiviral vectors. *Viruses* **3**, 132–159 (2011).

117. Chen, Y. H., Keiser, M. S. & Davidson, B. L. Viral vectors for gene transfer. *Curr. Protoc. Mouse Biol.* **8**, e58 (2018).

118. Yang, Y., Li, Q., Ertl, H. C. & Wilson, J. M. Cellular and humoral immune responses to viral antigens create barriers to lung-directed gene therapy with recombinant adenoviruses. *J. Virol.* **69**, 2004–2015 (1995).

119. Crystal, R. G. *et al.* Administration of an adenovirus containing the human CFTR cDNA to the respiratory tract of individuals with cystic fibrosis. *Nat. Genet.* **8**, 42–51 (1994).

120. Brunetti-Pierri, N. *et al.* Acute toxicity after high-dose systemic injection of helper-dependent adenoviral vectors into nonhuman primates. *Hum. Gene Ther.* **15**, 35–46 (2004).

121. Wilson, J. M. Lessons learned from the gene therapy trial for ornithine transcarbamylase deficiency. *Mol. Genet. Metab.* **96**, 490–497 (2009).

122. Brunetti-Pierri, N. & Ng, P. Gene therapy with helper-dependent adenoviral vectors: lessons from studies in large animal models. *Virus Genes* **53**, 684–691 (2017).

123. Zhang, J. *et al.* A novel oncolytic adenovirus targeting Wnt signaling effectively

inhibits cancer-stem like cell growth via metastasis, apoptosis and autophagy in HCC models. *Biochem. Biophys. Res. Commun.* **491**, 469–477 (2017).

124. Li, Y. *et al.* Anti-cancer effect of oncolytic adenovirus-armed shRNA targeting MYCN gene on doxorubicin-resistant neuroblastoma cells. *Biochem. Biophys. Res. Commun.* **491**, 134–139 (2017).

125. Joyce, C. *et al.* Orally administered adenoviral-based vaccine induces respiratory mucosal memory and protection against RSV infection in cotton rats. *Vaccine* **36**, 4265–4277 (2018).

126. Peng, Z. Current status of gendicine in China: Recombinant human Ad-p53 agent for treatment of cancers. *Hum. Gene Ther.* **16**, 1016–1027 (2005).

127. Liang, M. Oncorine, the world first oncolytic virus medicine and its update in China. *Curr. Cancer Drug Targets* **18**, 171–176 (2018).

128. Manservigi, R., Argnani, R. & Marconi, P. HSV recombinant vectors for gene therapy. *Open Virol. J.* **4**, 123–156 (2010).

129. Gambini, E. *et al.* Replication-competent herpes simplex virus retargeted to HER2 as therapy for high-grade glioma. *Mol. Ther.* **20**, 994–1001 (2012).

130. Antoszczyk, S. *et al.* Treatment of orthotopic malignant peripheral nerve sheath tumors with oncolytic herpes simplex virus. *Neuro. Oncol.* **16**, 1057–1066 (2014).

131. Kuruppu, D. & Tanabe, K. K. HSV-1 as a novel therapy for breast cancer meningeal metastases. *Cancer Gene Ther.* **22**, 506–508 (2015).

132. Waters, A. M. *et al.* Rationale and design of a phase 1 clinical trial to evaluate HSV G207 alone or with a single radiation dose in children with progressive or recurrent malignant supratentorial brain tumors. *Hum. Gene Ther. Clin. Dev.* **28**, 7–16 (2017).

133. Reach, T. FDA approves first oncolytic virus therapy: Imlygic for melanoma. *Oncol. Times* **37**, 36 (2015).

134. Daly, T. M. Overview of adeno-associated viral vectors. *Methods Mol. Biol.* **246**, 157–165 (2004).

135. Atchison, R. W., Casto, B. C. & Hammon, W. M. Adenovirus-associated defective virus particles. *Science.* **149**, 754–756 (1965).

136. Samulski, R. J., Berns, K. I., Tan, M. & Muzyczka, N. Cloning of adeno-associated virus into pBR322: Rescue of intact virus from the recombinant plasmid in human cells. *Proc. Natl. Acad. Sci. U. S. A.* **79**, 2077–2081 (1982).

137. Srivastava, A., Lusby, E. W. & Berns, K. I. Nucleotide sequence and organization of the adeno-associated virus 2 genome. *J. Virol.* **45**, 555–564 (1982).

138. Laughlin, C. A., Westphal, H. & Carter, B. J. Spliced adenovirus-associated virus RNA. *Proc. Natl. Acad. Sci. U. S. A.* **76**, 5567–5571 (1979).

139. Samulski, R. J. & Muzyczka, N. AAV-mediated gene therapy for research and therapeutic purposes. *Annu. Rev. Virol.* **1**, 427–451 (2014).

140. Lusby, E., Fife, K. H. & Berns, K. I. Nucleotide sequence of the inverted terminal repetition in adeno-associated virus DNA. *J. Virol.* **34**, 402–409 (1980).

141. McLaughlin, S. K., Collis, P., Hermonat, P. L. & Muzyczka, N. Adeno-associated virus general transduction vectors: analysis of proviral structures. *J. Virol.* **62**, 1963–1973 (1988).

142. Pereira, D. J., McCarty, D. M. & Muzyczka, N. The adeno-associated virus (AAV) Rep protein acts as both a repressor and an activator to regulate AAV transcription during a productive infection. *J. Virol.* **71**, 1079–1088 (1996).

143. Sonntag, F., Schmidt, K. & Kleinschmidt, J. A. A viral assembly factor promotes AAV2 capsid formation in the nucleolus. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 10220–10225 (2010).

144. Sonntag, F. *et al.* The assembly-activating protein promotes capsid assembly of different adeno-associated virus serotypes. *J. Virol.* **85**, 12686–12697 (2011).

145. Naumer, M. *et al.* Properties of the adeno-associated virus assembly-activating protein. *J. Virol.* **86**, 13038–13048 (2012).

146. Earley, L. F., Powers, J. M., Adachi, K. & Baumgart, J. T. Adeno-associated virus (AAV) assembly-activating protein is not an essential requirement for capsid assembly of AAV serotypes 4, 5, and 11. *J. Virol.* **91**, 1–21 (2017).

147. Kronenberg, S., Kleinschmidt, J. A. & Böttcher, B. Electro cryo-microscopy and image reconstruction of adeno-associated virus type 2 empty capsids. *EMBO Rep.* **2**, 997–1002 (2001).

148. Fitzpatrick, Z. *et al.* Influence of pre-existing anti-capsid neutralizing and binding antibodies on AAV vector transduction. *Mol. Ther. - Methods Clin. Dev.* **9**, 119–129 (2018).

149. Hermonat, P. L., Labow, M. A., Wright, R., Berns, K. I. & Muzyczka, N. Genetics of adeno-associated virus: isolation and preliminary characterization of adeno-associated virus type 2 mutants. *J. Virol.* **51**, 329–339 (1984).

150. Hoque, M. *et al.* Nuclear transport of the major capsid protein is essential for adeno-associated virus capsid formation. *J. Virol.* **73**, 7912–7915 (1999).

151. Naso, M. F., Tomkowicz, B., Perry, W. L. & Strohl, W. R. Adeno-associated virus (AAV) as a vector for gene therapy. *BioDrugs* **31**, 317–334 (2017).

152. Mak, K. Y., Rajapaksha, I. G., Angus, P. W. & Herath, C. B. The Adeno-associated Virus - A safe and promising vehicle for liver-specific gene therapy of inherited and non-inherited disorders. *Curr. Gene Ther.* **17**, (2017).

153. Pillay, S. *et al.* An essential receptor for adeno-associated virus infection. *Nature* **530**, 108–112 (2016).

154. Duan, D. *et al.* Dynamin is required for recombinant adeno-associated virus type 2 infection. *J. Virol.* **73**, 10371–10376 (1999).

155. Bartlett, J. S., Wilcher, R. & Samulski, R. J. Infectious entry pathway of adeno-associated virus and adeno-associated virus vectors. *J. Virol.* **74**, 2777–2785 (2000).

156. Douar, A., Poulard, K., Stockholm, D. & Danos, O. Intracellular trafficking of adeno-associated virus vectors: Routing to the late endosomal compartment and proteasome degradation. *J. Virol.* **75**, 1824–1833 (2001).

157. Nonnenmacher, M. & Weber, T. Adeno-associated virus 2 infection requires endocytosis through the CLIC/GEEC pathway. *Cell Host Microbe* **10**, 563–576 (2011).

158. Weinberg, M. S. *et al.* Recombinant adeno-associated virus utilizes cell-specific infectious entry mechanisms. *J. Virol.* **88**, 12472–12484 (2014).

159. Nonnenmacher, M. E., Cintrat, J.-C., Gillet, D. & Weber, T. Syntaxin 5-dependent retrograde transport to the trans-Golgi network is required for adeno-associated virus transduction. *J. Virol.* **89**, 1673–1687 (2015).

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

161. Grieger, J. C., Johnson, J. S., Gurda-Whitaker, B., Agbandje-McKenna, M. & Samulski, R. J. Surface-exposed Adeno-Associated Virus Vp1-NLS capsid fusion protein rescues infectivity of noninfectious Wild-Type Vp2/Vp3 and Vp3-only capsids but not that of fivefold pore mutant virions. *J. Virol.* **81**, 7833–7843 (2007).

162. Sonntag, F., Bleker, S., Leuchs, B., Fischer, R. & Kleinschmidt, J. A. Adeno-associated virus type 2 capsids with externalized VP1/VP2 trafficking domains are generated prior to passage through the cytoplasm and are maintained until uncoating occurs in the nucleus. *J. Virol.* **80**, 11040–11054 (2006).

163. Johnson, J. S. & Samulski, R. J. Enhancement of adeno-associated virus infection by mobilizing capsids into and out of the nucleolus. *J. Virol.* **83**, 2632–2644 (2009).

164. Li, C. *et al.* Adeno-associated virus capsid antigen presentation is dependent on endosomal escape. *J. Clin. Invest.* **123**, 1390–1401 (2013).

165. Xiao, P. J., Li, C., Neumann, A. & Samulski, R. J. Quantitative 3D tracing of gene-delivery viral vectors in human cells and animal tissues. *Mol. Ther.* **20**, 317–328 (2012).

166. Ferrari, F. K., Samulski, T., Shenk, T. & Samulski, R. J. Second-strand synthesis is a rate-limiting step for efficient transduction by recombinant adeno-associated virus vectors. *J. Virol.* **70**, 3227–3234 (1996).

167. Nakai, H., Storm, T. A. & Kay, M. A. Recruitment of single-stranded recombinant adeno-associated virus vector genomes and intermolecular recombination are responsible for stable transduction of liver in vivo. *J. Virol.* **74**, 9451–9463 (2000).

168. Kotin, R. M. *et al.* Site-specific integration by adeno-associated virus. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 2211–2215 (1990).

169. Nakai, H. *et al.* Extrachromosomal recombinant adeno-associated virus vector genomes are primarily responsible for stable liver transduction in vivo. *J. Virol.* **75**, 6969–6976 (2001).

170. Li, C. & Samulski, R. J. Engineering adeno-associated virus vectors for gene therapy. *Nat. Rev. Genet.* **21**, 255–272 (2020).

171. Afione, S. A. *et al.* In vivo model of adeno-associated virus vector persistence and rescue. *J. Virol.* **70**, 3235–3241 (1996).

172. Snyder, R. O. *et al.* Features of the adeno-associated virus origin involved in

substrate recognition by the viral Rep protein. *J. Virol.* **67**, 6096–6104 (1993).

173. Brister, J. R. & Muzyczka, N. Mechanism of Rep-mediated adeno-associated virus origin nicking. *J. Virol.* **74**, 7762–7771 (2000).

174. Myers, M. W. & Carter, B. J. Adeno-associated virus replication. *J. Biol. Chem.* **256**, 567–570 (1980).

175. Dubielzig, R., King, J. A., Weger, S., Kern, A. & Kleinschmidt, J. A. Adeno-associated virus type 2 protein interactions: Formation of pre-encapsidation complexes. *J. Virol.* **73**, 8989–8998 (1999).

176. King, J. A., Dubielzig, R., Grimm, D. & Kleinschmidt, J. A. DNA helicase-mediated packaging of adeno-associated virus type 2 genomes into preformed capsids. *EMBO J.* **20**, 3282–3291 (2001).

177. Colella, P., Ronzitti, G. & Mingozzi, F. Emerging issues in AAV-mediated in vivo gene therapy. *Mol. Ther. - Methods Clin. Dev.* **8**, 87–104 (2018).

178. Hurlbut, G. D. *et al.* Preexisting immunity and low expression in primates highlight translational challenges for liver-directed AAV8-mediated gene therapy. *Mol. Ther.* **18**, 1983–1994 (2010).

179. Grieger, J. C. & Samulski, R. J. Adeno-associated virus vectorology, manufacturing, and clinical applications. *Methods in Enzymology*. **507**, 229–254 (2012).

180. Grimm, D., Pandey, K., Nakai, H., Storm, T. A. & Kay, M. A. Liver transduction with recombinant adeno-associated virus is primarily restricted by capsid serotype not vector genotype. *J. Virol.* **80**, 426–439 (2006).

181. Palaschak, B., Herzog, R. W. & Markusic, D. M. AAV-mediated gene delivery to the liver: Overview of current technologies and methods. *Methods Mol. Biol.* **1950**, 333–360 (2019).

182. Zolotukhin, S. Production of recombinant adeno-associated virus vectors. *Hum. Gene Ther.* **557**, 551–557 (2005).

183. Wang, L., Blouin, V., Brument, N., Bello-Roufai, M. & Francois, A. Production and purification of recombinant adeno-associated vectors. *Methods in Molecular Biology*. **807**, 361–404 (2011).

184. Xiao, X., Li, J. & Samulski, R. J. Production of high-titer recombinant adeno-associated virus vectors in the absence of helper adenovirus. *J. Virol.* **72**, 2224–2232 (1998).

185. Conway, J. E. *et al.* High-titer recombinant adeno-associated virus production utilizing a recombinant herpes simplex virus type I vector expressing AAV-2 rep and cap. *Gene Ther.* **6**, 986–993 (1999).

186. Kang, W. *et al.* An efficient rHSV-based complementation system for the production of multiple rAAV vector serotypes. *Gene Ther.* **16**, 229–239 (2009).

187. Urabe, M., Ding, C. & Kotin, R. M. Insect cells as a factory to produce adeno-associated virus type 2 vectors. *Hum. Gene Ther.* **13**, 1935–1943 (2002).

188. Zolotukhin, S. *et al.* Recombinant adeno-associated virus purification using novel methods improves infectious titer and yield. *Gene Ther.* **6**, 973–985 (1999).

189. Gao, G. *et al.* Clades of adeno-associated viruses are widely disseminated in human tissues. *J. Virol.* **78**, 6381–6388 (2004).

190. Schmidt, M. *et al.* Adeno-associated virus type 12 (AAV12): a novel AAV serotype with sialic acid- and heparan sulfate proteoglycan-independent transduction activity. *J. Virol.* **82**, 1399–1406 (2008).

191. Zincarelli, C., Soltys, S., Rengo, G. & Rabinowitz, J. E. Analysis of AAV serotypes 1-9 mediated gene expression and tropism in mice after systemic injection. *Mol. Ther.* **16**, 1073–1080 (2008).

192. Wu, P. *et al.* Mutational analysis of the adeno-associated virus type 2 (AAV2) capsid gene and construction of AAV2 vectors with altered tropism. *J. Virol.* **74**, 8635–8647 (2000).

193. Kotterman, M. A., Chalberg, T. W. & Schaffer, D. V. Viral vectors for gene therapy: translational and clinical outlook. *Annu. Rev. Biomed. Eng.* **17**, 63–89 (2015).

194. Plotkin, J. B., Robins, H. & Levine, A. J. Tissue-specific codon usage and the expression of human genes. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 12588–12591 (2004).

195. Kattenhorn, L. M. *et al.* Adeno-associated virus gene therapy for liver disease. *Hum. Gene Ther.* **27**, 947–961 (2016).

196. Zaiss, A. *et al.* Differential activation of innate immune responses by adenovirus and adeno-associated virus vectors. *J. Virol.* **76**, 4580–4590 (2002).

197. Miller, D. G. *et al.* Large-scale analysis of adeno-associated virus vector integration sites in normal human cells. *J. Virol.* **79**, 11434–11442 (2005).

198. Kaeppel, C. *et al.* A largely random AAV integration profile after LPLD gene therapy. *Nat. Med.* **19**, 889–891 (2013).

199. Mingozzi, F. *et al.* Induction of immune tolerance to coagulation factor IX antigen by in vivo hepatic gene transfer. *J. Clin. Invest.* **111**, 1347–1356 (2003).

200. Dobrzynski, E. *et al.* Induction of antigen-specific CD4+ T-cell anergy and deletion by in vivo viral gene transfer. *Blood* **104**, 969–977 (2004).

201. Dobrzynski, E. *et al.* Prevention of cytotoxic T lymphocyte responses to factor IX-expressing hepatocytes by gene transfer-induced regulatory T cells. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 4592–4597 (2006).

202. Cooper, M. *et al.* Improved induction of immune tolerance to factor IX by hepatic AAV-8 gene transfer. *Hum. Gene Ther.* **20**, 767–776 (2009).

203. Martino, A. T. *et al.* Tolerance induction to cytoplasmic β-galactosidase by hepatic AAV gene transfer - Implications for antigen presentation and immunotoxicity. *PLoS One* **4**, (2009).

204. LoDuca, P., Hoffman, B. & Herzog, R. Hepatic gene transfer as a means of tolerance induction to transgene products. *Curr. Gene Ther.* **9**, 104–114 (2009).

205. Nathwani, A. C. *et al.* Long-term safety and efficacy of factor IX gene therapy in hemophilia B. *N. Engl. J. Med.* **371**, 1994–2004 (2014).

206. Kumar, S. R. P., Hoffman, B. E., Terhorst, C., de Jong, Y. P. & Herzog, R. W. The balance between CD8+ T cell-mediated clearance of AAV-encoded antigen in the

liver and tolerance is dependent on the vector dose. *Mol. Ther.* **25**, 880–891 (2017).

207. Keeler, G. D., Markusic, D. M. & Hoffman, B. E. Liver induced transgene tolerance with AAV vectors. *Cell. Immunol.* **342**, (2019).

208. Dane, A. P., Wowro, S. J., Cunningham, S. C. & Alexander, I. E. Comparison of gene transfer to the murine liver following intraperitoneal and intraportal delivery of hepatotropic AAV pseudo-serotypes. *Gene Ther.* **20**, 460–464 (2013).

209. Vercauteren, K. *et al.* Superior in vivo transduction of human hepatocytes using engineered AAV3 capsid. *Mol. Ther.* **24**, 1042–1049 (2016).

210. Lisowski, L. *et al.* Selection and evaluation of clinically relevant AAV variants in a xenograft liver model. *Nature* **506**, 382–386 (2014).

211. Wang, L. *et al.* Comparative study of liver gene transfer with AAV vectors based on natural and engineered AAV capsids. *Mol. Ther.* **23**, 1877–1887 (2015).

212. Nakai, H. *et al.* Unrestricted hepatocyte transduction with adeno-associated virus serotype 8 vectors in mice. *J. Virol.* **79**, 214–224 (2005).

213. Davidoff, A. M., Ng, C. Y. C., Zhou, J., Spence, Y. & Nathwani, A. C. Sex significantly influences transduction of murine liver by recombinant adeno-associated viral vectors through an androgen-dependent pathway. *Blood* **102**, 480–488 (2003).

214. Sarkar, R. *et al.* Total correction of hemophilia A mice with canine FVIII using an AAV 8 serotype. *Blood* **103**, 1253–1260 (2004).

215. Jiang, H. *et al.* Multiyear therapeutic benefit of AAV serotypes 2, 6, and 8 delivering factor VIII to hemophilia A mice and dogs. *Blood* **108**, 107–115 (2006).

216. Wang, L. *et al.* Sustained correction of disease in naive and AAV2-pretreated hemophilia B dogs: AAV2/8-mediated, liver-directed gene therapy. *Blood* **105**, 3079–3086 (2005).

217. Barbon, C. M. *et al.* AAV8-mediated hepatic expression of acid sphingomyelinase corrects the metabolic defect in the visceral organs of a mouse model of Niemann-Pick disease. *Mol. Ther.* **12**, 431–440 (2005).

218. Sun, B. *et al.* Efficacy of an adeno-associated virus 8-pseudotyped vector in glycogen storage disease type II. *Mol. Ther.* **11**, 57–65 (2005).

219. Koeberl, D. D. *et al.* Early, sustained efficacy of adeno-associated virus vector-mediated gene therapy in glycogen storage disease type Ia. *Gene Ther.* **13**, 1281–1289 (2006).

220. Lebherz, C. *et al.* Gene therapy with novel adeno-associated virus vectors substantially diminishes atherosclerosis in a murine model of familial hypercholesterolemia. *J. Gene Med.* **6**, 663–672 (2004).

221. Kitajima, K. *et al.* Complete prevention of atherosclerosis in ApoE-deficient mice by hepatic human ApoE gene transfer with adeno-associated virus serotypes 7 and 8. *Arterioscler. Thromb. Vasc. Biol.* **26**, 1852–1857 (2006).

222. Berns, K. I. & Muzyczka, N. AAV: An overview of unanswered questions. *Hum. Gene Ther.* **28**, 308–313 (2017).

223. Dong, B., Nakai, H. & Xiao, W. Characterization of genome integrity for oversized

recombinant AAV vector. *Mol. Ther.* **18**, 87–92 (2010).

224. Wu, Z., Yang, H. & Colosi, P. Effect of genome size on AAV vector packaging. *Mol. Ther.* **18**, 80–86 (2010).

225. Burton, M. *et al.* Coexpression of factor VIII heavy and light chain adeno-associated viral vectors produces biologically active protein. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 12725–12730 (1999).

226. Duan, D., Yue, Y. & Engelhardt, J. F. Expanding AAV packaging capacity with trans-splicing or overlapping vectors: A quantitative comparison. *Mol. Ther.* **4**, 383–391 (2001).

227. McCarty, D. M., Monahan, P. E. & Samulski, R. J. Self-complementary recombinant adeno-associated virus (scAAV) vectors promote efficient transduction independently of DNA synthesis. *Gene Ther.* **8**, 1248–1254 (2001).

228. Russell, D. W. & Hirata, R. K. Human gene targeting by viral vectors. *Nat. Genet.* **18**, 231–236 (1998).

229. Chandler, R. J. *et al.* Vector design influences hepatic genotoxicity after adeno-associated virus gene therapy. *J. Clin. Invest.* **125**, 870–880 (2015).

230. Moscioni, D. *et al.* Long-term correction of ammonia metabolism and prolonged survival in ornithine transcarbamylase-deficient mice following liver-directed treatment with adeno-associated viral vectors. *Mol. Ther.* **14**, 25–33 (2006).

231. Bell, P. *et al.* No evidence for tumorigenesis of AAV vectors in a large-scale study in mice. *Mol. Ther.* **12**, 299–306 (2005).

232. Li, H., Malani, N. & Hamilton, S. R. Assessing the potential for AAV vector genotoxicity in a murine model. *Blood* **117**, 6739 (2011).

233. Nathwani, A. C. *et al.* Long-term safety and efficacy following systemic administration of a self-complementary AAV vector encoding human FIX pseudotyped with serotype 5 and 8 capsid proteins. *Mol. Ther.* **19**, 876–885 (2011).

234. Gil-Farina, I. *et al.* Recombinant AAV integration is not associated with hepatic genotoxicity in nonhuman primates and patients. *Mol. Ther.* **24**, 1100–1105 (2016).

235. Russell, D. W. & Grompe, M. Adeno-associated virus finds its disease. *Nat. Genet.* **47**, 1104–1105 (2015).

236. Nault, J. C. *et al.* Recurrent AAV2-related insertional mutagenesis in human hepatocellular carcinomas. *Nat. Genet.* **47**, 1187–1193 (2015).

237. Berns, K. I. *et al.* Adeno-associated virus type 2 and hepatocellular carcinoma? *Hum. Gene Ther.* **26**, 779–781 (2015).

238. Nault, J. C. *et al.* Wild-type AAV insertions in hepatocellular carcinoma do not inform debate over genotoxicity risk of vectorized AAV. *Mol. Ther.* **24**, 660–661 (2016).

239. Manno, C. S. *et al.* Successful transduction of liver in hemophilia by AAV-Factor IX and limitations imposed by the host immune response. *Nat. Med.* **12**, 342–347 (2006).

240. Favre, D. *et al.* Immediate and long-term safety on recombinant adeno-associated

virus injection into the nonhuman primate muscle. *Mol. Ther.* **4**, 559–566 (2001).

241. Arruda, V. R. *et al.* Lack of germline transmission of vector sequences following systemic administration of recombinant AAV-2 vector in males. *Mol. Ther.* **4**, 586–592 (2001).

242. Schuettrumpf, J. *et al.* Inadvertent germline transmission of AAV2 vector: Findings in a rabbit model correlate with those in a human clinical trial. *Mol. Ther.* **13**, 1064–1073 (2006).

243. Vandamme, C., Adjali, O. & Mingozzi, F. Unraveling the complex story of immune responses to AAV vectors trial after trial. *Hum. Gene Ther.* **28**, 1061–1074 (2017).

244. Basner-Tschakarjan, E. & Mingozzi, F. Cell-mediated immunity to AAV vectors, evolving concepts and potential solutions. *Front. Immunol.* **5**, 1–10 (2014).

245. Rogers, G. L. *et al.* Innate immune responses to AAV vectors. *Front. Microbiol.* **2**, 1–10 (2011).

246. Hösel, M. *et al.* Toll-like receptor 2-mediated innate immune response in human nonparenchymal liver cells toward adeno-associated viral vectors. *Hepatology* **55**, 287–297 (2012).

247. Kawai, T. & Akira, S. Signaling to NF-κB by Toll-like receptors. *Trends Mol. Med.* **13**, 460–469 (2007).

248. Zhu, J., Huang, X. & Yang, Y. The TLR9-MyD88 pathway is critical for adaptive immune responses to adenoassociated virus gene therapy vectors in mice. *J. Clin. Invest.* **119**, 2388–2398 (2009).

249. Shao, W. *et al.* Double-stranded RNA innate immune response activation from long-term adeno-associated virus vector transduction. *JCI insight* **3**, (2018).

250. Nidetz, N. F. *et al.* Adeno-associated viral vector-mediated immune responses: Understanding barriers to gene delivery. *Pharmacol. Ther.* **207**, 107453 (2020).

251. Faust, S. M. *et al.* CpG-depleted adeno-associated virus vectors evade immune detection Find the latest version : CpG-depleted adeno-associated virus vectors evade immune detection. *J. Clin. Invest.* **123**, 2994–3001 (2013).

252. Mingozzi, F. *et al.* Overcoming preexisting humoral immunity to AAV using capsid decoys. *Sci. Transl. Med.* **5**, (2013).

253. Calcedo, R., Vandenberghe, L. H., Gao, G., Lin, J. & Wilson, J. M. Worldwide epidemiology of neutralizing antibodies to adeno-associated viruses. *J. Infect. Dis.* **199**, 381–390 (2009).

254. Boutin, S., Monteilhet, V., Veron, P., Leborgne, C. & Benveniste, O. Prevalence of serum IgG and neutralizing factors and 9 in the healthy population: Implications for gene therapy using AAV vectors. *Hum. Gene Ther.* **21**, 704–712 (2010).

255. Calcedo, R. *et al.* Pre-existing neutralizing antibodies to AAV capsids in large animals other than monkeys may confound in vivo gene therapy studies Roberto. *Hum. Gene Ther. Methods* **26**, 103–105 (2015).

256. Li, C. *et al.* Single amino acid modification of adeno-associated virus capsid changes transduction and humoral immune profiles. *J. Virol.* **86**, 7752–7759 (2012).

257. Calcedo, R. *et al.* Adeno-associated virus antibody profiles in newborns, children, and adolescents. *Clin. Vaccine Immunol.* **18**, 1586–1588 (2011).

258. Majowicz, A. *et al.* Successful repeated hepatic gene delivery in mice and non-human primates achieved by sequential administration of AAV5ch and AAV1. *Mol. Ther.* **25**, 1831–1842 (2017).

259. Ertl, H. C. J. & High, K. A. Impact of AAV capsid-specific T-cell responses on design and outcome of clinical gene transfer trials with recombinant adeno-associated viral vectors: An evolving controversy. *Hum. Gene Ther.* **28**, 328–337 (2017).

260. Maersch, S., Huber, A., Büning, H., Hallek, M. & Perabo, L. Optimization of stealth adeno-associated virus vectors by randomization of immunogenic epitopes. *Virology* **397**, 167–175 (2010).

261. Bartel, M., Schaffer, D. & Büning, H. Enhancing the clinical potential of AAV vectors by capsid engineering to evade pre-existing immunity. *Front. Microbiol.* **2**, 1–10 (2011).

262. Tse, L. V. *et al.* Structure-guided evolution of antigenically distinct adeno-associated virus variants for immune evasion. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E4812–E4821 (2017).

263. Meliani, A. *et al.* Enhanced liver gene transfer and evasion of preexisting humoral immunity with exosome-enveloped AAV vectors. *Blood Adv.* **1**, 2019–2031 (2017).

264. György, B. & Maguire, C. A. Extracellular vesicles: nature's nanoparticles for improving gene transfer with adeno-associated virus vectors. *Wiley Interdiscip. Rev. Nanomedicine Nanobiotechnology* **10**, 1–13 (2018).

265. Monteilhet, V. *et al.* A 10 patient case report on the impact of plasmapheresis upon neutralizing factors against adeno-associated virus (AAV) types 1, 2, 6, and 8. *Mol. Ther.* **19**, 2084–2091 (2011).

266. Mimuro, J. *et al.* Minimizing the inhibitory effect of neutralizing antibody for efficient gene expression in the liver with adeno-associated virus 8 vectors. *Mol. Ther.* **21**, 318–323 (2013).

267. Mays, L. E. *et al.* Adeno-associated virus capsid structure drives CD4-dependent CD8 + T cell response to vector encoded proteins. *J. Immunol.* **182**, 6051–6060 (2009).

268. Mays, L. E. *et al.* AAV8 induces tolerance in murine muscle as a result of poor APC transduction, T cell exhaustion, and minimal MHCI upregulation on target cells. *Mol. Ther.* **22**, 28–41 (2014).

269. Mingozzi, F. *et al.* CD18+ T-cell responses to adeno-associated virus capsid in humans. *Nat. Med.* **13**, 419–422 (2007).

270. Nathwani, A. C. *et al.* Adenovirus-associated virus vector-mediated gene transfer in hemophilia B. *N. Engl. J. Med.* **365**, 2357–2365 (2011).

271. George, L. A. *et al.* Hemophilia B gene therapy with a high-specific-activity factor IX variant. *N. Engl. J. Med.* **377**, 2215–2227 (2017).

272. Duan, D., Yue, Y., Yan, Z., Yang, J. & Engelhardt, J. F. Endosomal processing limits gene transfer to polarized airway epithelia by adeno-associated virus. *J. Clin. Invest.*

105, 1573–1587 (2000).

273. Finn, J. D. *et al.* Proteasome inhibitors decrease AAV2 capsid derived peptide epitope presentation on mhc class i following transduction. *Mol. Ther.* **18**, 135–142 (2010).

274. Martino, A. T. *et al.* Engineered AAV vector minimizes in vivo targeting of transduced hepatocytes by capsid-specific CD8+ T cells. *Blood* **121**, 2224–2233 (2013).

275. Lu, Y. & Song, S. Distinct immune responses to transgene products from rAAV1 and rAAV8 vectors. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 17158–17162 (2009).

276. Finn, J. D. *et al.* Eradication of neutralizing antibodies to factor VIII In canine hemophilia a after liver gene therapy. *Blood* **116**, 5842–5848 (2010).

277. Crispe, I. N. Immune tolerance in liver disease. *Hepatology* **60**, 2109–2117 (2014).

278. Mingozzi, F. & High, K. A. Therapeutic in vivo gene transfer for genetic disease using AAV: Progress and challenges. *Nat. Rev. Genet.* **12**, 341–355 (2011).

279. Gaudet, D. *et al.* Efficacy and long-term safety of alipogene tiparvovec (AAV1-LPL S447X) gene therapy for lipoprotein lipase deficiency: An open-label trial. *Gene Ther.* **20**, 361–369 (2013).

280. Russell, S. *et al.* Efficacy and safety of voretigene neparvovec (AAV2-hRPE65v2) in patients with RPE65-mediated inherited retinal dystrophy: a randomised, controlled, open-label, phase 3 trial. *Lancet* **390**, 849–860 (2017).

281. Rangarajan, S. *et al.* AAV5-factor VIII gene transfer in severe hemophilia a. *N. Engl. J. Med.* **377**, 2519–2530 (2017).

282. Malone, D. C. *et al.* Cost-effectiveness analysis of using onasemnogene abeparvocec (AVXS-101) in spinal muscular atrophy type 1 patients. *J. Mark. Access Heal. Policy* **7**, 1601484 (2019).

283. Pasi, K. J. *et al.* Multiyear follow-up of AAV5-hfVIII-sq gene therapy for hemophilia A. *N. Engl. J. Med.* **382**, 29–40 (2020).

284. Flotte, T. *et al.* A phase I study of an adeno-associated Virus-CFTR gene vector in adult CF patients with mild lung disease. *Hum. Gene Ther.* **7**, 1145–1159 (1996).

285. Dunbar, C. E. *et al.* Gene therapy comes of age. *Science.* **359**, (2018).

286. Bryant, L. M. *et al.* Lessons learned from the clinical development and market authorization of Glybera. *Hum. Gene Ther. Clin. Dev.* **24**, 55–64 (2013).

287. Mintzer, M. A. & Simanek, E. E. Nonviral vectors for gene delivery. *Chem. Rev.* **109**, 259–302 (2009).

288. Yin, H. *et al.* Non-viral vectors for gene-based therapy. *Nat. Rev. Genet.* **15**, 541–555 (2014).

289. Kowalski, P. S., Rudra, A., Miao, L. & Anderson, D. G. Delivering the messenger: Advances in technologies for therapeutic mRNA delivery. *Mol. Ther.* **27**, 710–728 (2019).

290. Fornaguera, C., Castells-Sala, C. & Borrós, S. Unraveling polymeric nanoparticles cell

uptake pathways: Two decades working to understand nanoparticles journey to improve gene therapy. *Adv Exp Med Biol* **1288**, 117–138 (2020).

291. Segovia, N. *et al.* Hydrogel doped with nanoparticles for local sustained release of siRNA in breast cancer. *Adv. Healthc. Mater.* **4**, 271–280 (2015).

292. Dosta, P., Segovia, N., Cascante, A., Ramos, V. & Borrós, S. Surface charge tunability as a powerful strategy to control electrostatic interaction for high efficiency silencing, using tailored oligopeptide-modified poly(beta-amino ester)s (PBAEs). *Acta Biomater.* **20**, 82–93 (2015).

293. Fornaguera, C. *et al.* mRNA delivery system for targeting antigen-presenting cells in vivo. *Adv. Healthc. Mater.* **7**, 1–11 (2018).

294. Fornaguera, C. *et al.* In vivo retargeting of poly(beta aminoester) (OM-PBAE) nanoparticles is influenced by protein corona. *Adv. Healthc. Mater.* **8**, 1–11 (2019).

295. Riera, R. *et al.* Tracking the DNA complexation state of pBAE polyplexes in cells with super resolution microscopy. *Nanoscale* **11**, 17869–17877 (2019).

296. Balcells, L. *et al.* SPIONs' enhancer effect on cell transfection: An unexpected advantage for an improved gene delivery system. *ACS Omega* **4**, 2728–2740 (2019).

297. Brugada-Vilà, P. *et al.* Oligopeptide-modified poly(beta-amino ester)s-coated AdNuPARmE1A: Boosting the efficacy of intravenously administered therapeutic adenoviruses. *Theranostics* **10**, 2744–2758 (2020).

298. Kulkarni, J. A., Cullis, P. R. & Van Der Meel, R. Lipid nanoparticles enabling gene therapies: From concepts to clinical utility. *Nucleic Acid Ther.* **28**, 146–157 (2018).

299. Chen, F., Alphonse, M. & Liu, Q. Strategies for nonviral nanoparticle-based delivery of CRISPR/Cas9 therapeutics. *Wiley Interdiscip. Rev. Nanomedicine Nanobiotechnology* **9**, 1–14 (2019).

300. Liu, J., Zeng, F. & Allen, C. Influence of serum protein on polycarbonate-based copolymer micelles as a delivery system for a hydrophobic anti-cancer agent. *J. Control. Release* **103**, 481–497 (2005).

301. Jones, S. W. *et al.* Nanoparticle clearance is governed by Th1 / Th2 immunity and strain background find the latest version : Nanoparticle clearance is governed by Th1 / Th2 immunity and strain background. *J. Clin. Invest.* **123**, 3061–3073 (2013).

302. Suk, J. S., Xu, Q., Kim, N., Hanes, J. & Ensign, L. M. PEGylation as a strategy for improving nanoparticle-based drug and gene delivery. *Adv Funct Mater* **99**, 28–51 (2016).

303. Ando, H. *et al.* Reactivity of IgM antibodies elicited by PEGylated liposomes or PEGylated lipoplexes against auto and foreign antigens. *J. Control. Release* **270**, 114–119 (2018).

304. Yan, X. *et al.* The role of apolipoprotein E in the elimination of liposomes from blood by hepatocytes in the mouse. *Biochem. Biophys. Res. Commun.* **328**, 57–62 (2005).

305. Akinc, A. *et al.* Targeted delivery of RNAi therapeutics with endogenous and exogenous ligand-based mechanisms. *Mol. Ther.* **18**, 1357–1364 (2010).

306. Love, K. T. *et al.* Lipid-like materials for low-dose, in vivo gene silencing. *Proc. Natl.*

Acad. Sci. U. S. A. **107**, 1864–1869 (2010).

307. Jayaraman, M. *et al.* Maximizing the potency of siRNA lipid nanoparticles for hepatic gene silencing in vivo. *Angew. Chemie - Int. Ed.* **51**, 8529–8533 (2012).

308. Maier, M. A. *et al.* Biodegradable lipids enabling rapidly eliminated lipid nanoparticles for systemic delivery of RNAi therapeutics. *Mol. Ther.* **21**, 1570–1578 (2013).

309. Pardi, N. *et al.* Expression kinetics of nucleoside-modified mRNA delivered in lipid nanoparticles to mice by various routes Norbert. *J Control Release* **217**, 345–351 (2015).

310. Pardi, N. *et al.* Zika virus protection by a single low dose nucleoside modified mRNA vaccination. *Nature* **543**, 248–251 (2017).

311. Pardi, N. *et al.* Characterization of HIV-1 nucleoside-modified mRNA vaccines in rabbits and Rhesus macaques. *Mol. Ther. - Nucleic Acids* **15**, 36–47 (2019).

312. Conway, A. *et al.* Non-viral delivery of Zinc finger nuclease mRNA enables highly efficient in vivo genome editing of multiple therapeutic gene targets. *Mol. Ther.* **27**, 866–877 (2019).

313. Torres-Torronteras, J. *et al.* Hematopoietic gene therapy restores thymidine phosphorylase activity in a cell culture and a murine model of MNGIE. *Gene Ther.* **18**, 795–806 (2011).

314. Torres-Torronteras, J. *et al.* Gene therapy using a liver-targeted AAV vector restores nucleoside and nucleotide homeostasis in a murine model of MNGIE. *Mol. Ther.* **22**, 901–907 (2014).

315. Yadak, R. *et al.* Preclinical efficacy and safety evaluation of hematopoietic stem cell gene therapy in a mouse model of MNGIE. *Mol. Ther. - Methods Clin. Dev.* **8**, 152–165 (2018).

316. Torres-Torronteras, J. *et al.* Long-term restoration of thymidine phosphorylase function and nucleoside homeostasis using hematopoietic gene therapy in a murine model of Mitochondrial Neurogastrointestinal Encephalomyopathy. *Hum. Gene Ther.* **27**, 656–667 (2016).

317. Torres-Torronteras, J. *et al.* Long-term sustained effect of liver-targeted adeno-associated virus gene therapy for mitochondrial neurogastrointestinal encephalomyopathy. *Hum. Gene Ther.* **29**, 708–718 (2018).

318. Cabrera-Pérez, R. *et al.* Alpha-1-antitrypsin promoter improves the efficacy of an adeno-associated virus vector for the treatment of Mitochondrial Neurogastrointestinal Encephalomyopathy. *Hum. Gene Ther.* **30**, 985–998 (2019).

319. Shrivastav, M., De Haro, L. P. & Nickoloff, J. A. Regulation of DNA double-strand break repair pathway choice. *Cell Res.* **18**, 134–147 (2008).

320. Hartlerode, A. J. & Scully, R. Mechanisms of double-strand break repair in somatic mammalian cells. *Biochem. J.* **423**, 157–168 (2009).

321. Lieber, M. R., Gu, J., Lu, H., Shimazaki, N. & Tsai, A. G. Nonhomologous DNA end joining (NHEJ) and chromosomal translocations in humans. *Subcell. Biochem.* **50**, 279–296 (2010).

322. Chapman, J. R., Taylor, M. R. G. & Boulton, S. J. Playing the end game: DNA double-strand break repair pathway choice. *Mol. Cell* **47**, 497–510 (2012).

323. Capecchi, M. R. Altering the genome by homologous recombination. *Science.* **244**, 1288–1292 (1989).

324. Rouet, P., Smih, F. & Jasin, M. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol. Cell. Biol.* **14**, 8096–106 (1994).

325. Choulika, A., Perrin, A., Dujon, B. & Nicolas, J. F. Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of Saccharomyces cerevisiae. *Mol. Cell. Biol.* **15**, 1968–1973 (1995).

326. Liang, F., Han, M., Romanienko, P. J. & Jasin, M. Homology-directed repair is a major double-strand break repair pathway in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 5172–5177 (1998).

327. Paques, F. & Duchateau, P. Meganucleases and DNA double-strand break-induced recombination: Perspectives for gene therapy. *Curr. Gene Ther.* **7**, 49–66 (2007).

328. Delacôte, F. & Lopez, B. S. Importance of the cell cycle phase for the choice of the appropriate DSB repair pathway, for genome stability maintenance: The trans-S double-strand break repair model. *Cell Cycle* **7**, 33–38 (2008).

329. Li, H. *et al.* Applications of genome editing technology in the targeted therapy of human diseases: mechanisms, advances and prospects. *Signal Transduct. Target. Ther.* **5**, (2020).

330. Hirakawa, M. P., Krishnakumar, R., Timlin, J. A., Carney, J. P. & Butler, K. S. Gene editing and CRISPR in the clinic: current and future perspectives. *Biosci. Rep.* **40**, 1–37 (2020).

331. Pâques, F. & Haber, J. E. Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae. *Microbiol. Mol. Biol. Rev.* **63**, 349–404 (1999).

332. Chang, H. H. Y., Pannunzio, N. R., Adachi, N. & Lieber, M. R. Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Physiol. Behav.* **18**, 495–506 (2017).

333. Wang, H. & Xu, X. Microhomology-mediated end joining: New players join the team. *Cell Biosci.* **7**, 4–9 (2017).

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

335. Cathomen, T. & Keith Joung, J. Zinc-finger nucleases: The next generation emerges. *Mol. Ther.* **16**, 1200–1207 (2008).

336. Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S. & Gregory, P. D. Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.* **11**, 636–646 (2010).

337. Fairall, L., Schwabe, J. W. R., Chapman, L., Finch, J. T. & Rhodes, D. The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition. *Nature* **366**, 483–487 (1993).

338. Diakun, G. P., Fairall, L. & Klug, A. EXAFS study of the zinc-binding sites in the protein transcription factor IIIA. *Nature* **324**, 698–699 (1986).

339. Beerli, R. R., Schopfer, U., Dreier, B. & Barbas, C. F. Chemically regulated zinc finger transcription factors. *J. Biol. Chem.* **275**, 32617–32627 (2000).

340. Smith, J. Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. *Nucleic Acids Res.* **28**, 3361–3369 (2000).

341. Tebas, P. *et al.* Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. *N. Engl. J. Med.* **370**, 901–910 (2014).

342. Ashmore-Harris, C. & Fruhwirth, G. O. The clinical potential of gene editing as a tool to engineer cell-based therapeutics. *Clin. Transl. Med.* **9**, (2020).

343. Sharma, R. *et al.* In vivo genome editing of the albumin locus as a platform for protein replacement therapy. *Blood* **126**, 1777–1784 (2015).

344. Osborn, M. J. *et al.* Evaluation of TCR gene editing achieved by TALENs, CRISPR/Cas9, and megaTAL nucleases. *Mol. Ther.* **24**, 570–581 (2016).

345. Ou, L. *et al.* ZFN-mediated in vivo genome editing corrects murine Hurler syndrome. *Mol. Ther.* **27**, 178–187 (2019).

346. Laoharawee, K. *et al.* Dose-dependent prevention of metabolic and neurologic disease in murine MPS II by ZFN-mediated in vivo genome editing. *Mol. Ther.* **26**, 1127–1136 (2018).

347. Moscou, M. J. & Bogdanove, A. J. A simple cipher governs DNA recognition by TAL effectors. *Science.* **326**, 1501 (2009).

348. Boch, J. *et al.* Breaking the code of DNA binding specificity of TAL-type III effectors. *Science.* **326**, 1509–1512 (2009).

349. Christian, M. *et al.* Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* **186**, 756–761 (2010).

350. Li, T. *et al.* TAL nucleases (TALNs): Hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Res.* **39**, 359–372 (2011).

351. Zhang, F. *et al.* Programmable sequence-specific transcriptional regulation of mammalian genome using designer TAL effectors. *Nat. Biotechnol.* **29**, 149–153 (2011).

352. Qasim, W. *et al.* Molecular remission of infant B-ALL after infusion of universal TALEN gene-edited CAR T cells. *Sci. Transl. Med.* **9**, (2017).

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

354. Mojica, F. J. M., Díez-Villaseñor, C., Soria, E. & Juez, G. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol. Microbiol.* **36**, 244–246 (2000).

355. Jansen, R., Van Embden, J. D. A., Gaastra, W. & Schouls, L. M. Identification of genes

that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* **43**, 1565–1575 (2002).

356. Haft, D. H., Selengut, J., Mongodin, E. F. & Nelson, K. E. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/cas subtypes exist in prokaryotic genomes. *PLoS Comput. Biol.* **1**, 0474–0483 (2005).

357. Bolotin, A., Quinquis, B., Sorokin, A. & Dusko Ehrlich, S. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**, 2551–2561 (2005).

358. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **60**, 174–182 (2005).

359. Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I. & Koonin, E. V. A putative RNA-interference-based immune system in prokaryotes: Computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct* **1**, 1–26 (2006).

360. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science.* **315**, 1709–1712 (2007).

361. Brouns, S. J. J. *et al.* Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science.* **321**, 960–964 (2008).

362. Garneau, J. E. *et al.* The CRISPR/cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67–71 (2010).

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

364. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 2579–2586 (2012).

365. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. CRISPR-assisted editing of bacterial genomes. *Nat. Biotechnol.* **31**, 233–239 (2013).

366. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science.* **339**, 823–826 (2013).

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

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

369. Carroll, D. Genome engineering with targetable nucleases. *Annu. Rev. Biochem.* **83**, 409–439 (2014).

370. Gupta, D. *et al.* CRISPR-Cas9 system: A new-fangled dawn in gene editing. *Life Sci.* **232**, 116636 (2019).

371. Makarova, K. S. *et al.* Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* **18**, 67–83 (2020).

372. Moon, S. Bin, Kim, D. Y., Ko, J. H. & Kim, Y. S. Recent advances in the CRISPR genome

editing tool set. *Exp. Mol. Med.* **51**, (2019).

373. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733–740 (2009).

374. Horvath, P. & Barrangou, R. CRISPR/Cas, the immune system of Bacteria and Archaea. *Science.* **327**, 167–170 (2010).

375. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62–67 (2014).

376. Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935–959 (2014).

377. Palermo, G. *et al.* Key role of the REC lobe during CRISPR–Cas9 activation by 'sensing', 'regulating', and 'locking' the catalytic HNH domain. *Q Rev Biophys* **51**, (2018).

378. Zuo, Z. & Liu, J. Cas9-catalyzed DNA cleavage generates staggered ends: Evidence from molecular dynamics simulations. *Sci. Rep.* **5**, 1–9 (2016).

379. Zhang, S. *et al.* Recent advances of CRISPR/Cas9-based genetic engineering and transcriptional regulation in industrial biology. *Front. Bioeng. Biotechnol.* **7**, 1–11 (2020).

380. Deltcheva, E., Chylinski, K., Sharma, C. M. & Gonzales, K. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature.* **471**, 602–607 (2011).

381. Nuñez, J. K. *et al.* Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nat. Struct. Mol. Biol.* **21**, 528–534 (2014).

382. Wei, Y., Terns, R. M. & Terns, M. P. Cas9 function and host genome sampling in type II-A CRISPR–cas adaptation. *Genes Dev.* **29**, 356–361 (2015).

383. Heler, R. *et al.* Cas9 specifies functional viral targets during CRISPR-Cas adaptation. *Nature* **519**, 199–202 (2015).

384. Nuñez, J. K., Lee, A. S. Y., Engelman, A. & Doudna, J. A. Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature* **519**, 193–198 (2015).

385. Nuñez, J. K., Harrington, L. B., Kranzusch, P. J., Engelman, A. N. & Doudna, J. A. Foreign DNA capture during CRISPR-Cas adaptive immunity. *Nature* **527**, 535–538 (2015).

386. Sternberg, S. H., Richter, H., Charpentier, E. & Qimron, U. Adaptation in CRISPR-Cas systems. *Mol. Cell* **61**, 797–808 (2016).

387. Perez-Rodriguez, R. *et al.* Envelope stress is a trigger of CRISPR RNA-mediated DNA silencing in Escherichia coli. *Mol. Microbiol.* **79**, 584–599 (2011).

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

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

390. Zhang, X. H., Tee, L. Y., Wang, X. G., Huang, Q. S. & Yang, S. H. Off-target effects in CRISPR/Cas9-mediated genome engineering. *Mol. Ther. - Nucleic Acids* **4**, e264 (2015).

391. Carroll, D. Collateral damage: Benchmarking off-target effects in genome editing. *Genome Biol.* **20**, 19–21 (2019).

392. Tsai, S. Q. & Joung, J. K. Defining and improving the genome-wide specificities of CRISPR-Cas9 nucleases. *Nat. Rev. Genet.* **17**, 300–312 (2016).

393. Martin, F., Sánchez-Hernández, S., Gutiérrez-Guerrero, A., Pinedo-Gomez, J. & Benabdellah, K. Biased and unbiased methods for the detection of off-target cleavage by CRISPR/Cas9: An overview. *Int. J. Mol. Sci.* **17**, (2016).

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

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

396. Singh, K. *et al.* Efficient in vivo liver-directed gene editing using CRISPR/Cas9. *Mol. Ther.* **26**, 1241–1254 (2018).

397. Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.* **31**, 839–843 (2013).

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

399. Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science.* **351**, 84–88 (2016).

400. Kleinstiver, B. P. *et al.* Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. *Nat. Biotechnol.* **34**, 869–874 (2016).

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

402. Tsai, S. Q. *et al.* GUIDE-Seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, 187–197 (2015).

403. Kim, D. *et al.* Digenome-seq: Genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat. Methods* **12**, 237–243 (2015).

404. Crosetto, N. *et al.* Nucleotide-resolution DNA double-strand breaks mapping by next-generation sequencing. *Nat. Methods* **10**, 361–365 (2013).

405. Frock, R. L. *et al.* Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nat. Biotechnol.* **33**, 179–186 (2015).

406. Liu, G., Zhang, Y. & Zhang, T. Computational approaches for effective CRISPR guide RNA design and evaluation. *Comput. Struct. Biotechnol. J.* **18**, 35–44 (2020).

407. Kleinstiver, B. P. *et al.* High-fidelity CRISPR-Cas9 variants with undetectable genome- wide off-targets. *Nature* **528**, 490–495 (2016).

408. Chen, J. S. *et al.* Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* **550**, 407–410 (2017).

409. Vakulskas, C. A. *et al.* A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human haematopoietic stem and progenitor cells. *Nature Medicine*. **24**, 1216–1224 (2018).

410. Hyodo, T. *et al.* Tandem paired nicking promotes precise genome editing with scarce interference by p53. *Cell Rep.* **30**, 1195–1207 (2020).

411. Moreno-Mateos, M. A. *et al.* CRISPR-Cpf1 mediates efficient homology-directed repair and temperature-controlled genome editing. *Nat. Commun.* **8**, 1–9 (2017).

412. Sansbury, B. M., Wagner, A. M., Nitzan, E., Tarcic, G. & Kmiec, E. B. CRISPR-directed in vitro gene editing of plasmid DNA catalyzed by Cpf1 (Cas12a) nuclease and a mammalian cell-free extract. *Cris. J.* **1**, 191–202 (2018).

413. Finn, J. D. *et al.* A single administration of CRISPR/Cas9 lipid nanoparticles achieves robust and persistent in vivo genome editing. *Cell Rep.* **22**, 2227–2235 (2018).

414. Moon, S. Bin, Kim, D. Y., Ko, J. H., Kim, J. S. & Kim, Y. S. Improving CRISPR genome editing by engineering guide RNAs. *Trends Biotechnol.* **37**, 870–881 (2019).

415. Kim, S., Kim, D., Cho, S., Kim, J. & Kim, J.-S. Highly efficient RNA-guide genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.* **24**, 1012–1019 (2014).

416. Shin, J. *et al.* Disabling Cas9 by an anti-CRISPR DNA mimic. *Sci. Adv.* **3**, 1–9 (2017).

417. Nishiyama, J., Mikuni, T. & Yasuda, R. Virus-mediated genome editing via homology-directed repair in mitotic and postmitotic cells in mammalian brain. *Neuron* **96**, 755–768 (2017).

418. Yu, C. *et al.* Small molecules enhance CRISPR genome editing in pluripotent stem cells. *Cell Stem Cell* **16**, 142–147 (2015).

419. Pinder, J., Salsman, J. & Dellaire, G. Nuclear domain 'knock-in' screen for the evaluation and identification of small molecule enhancers of CRISPR-based genome editing. *Nucleic Acids Res.* **43**, 9379–9392 (2015).

420. Maruyama, T. *et al.* Inhibition of non-homologous end joining increases the efficiency of CRISPR/Cas9-mediated precise genome editing. *Nature* **33**, 538–542 (2015).

421. Chu, V. T. *et al.* Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat. Biotechnol.* **33**, 543–548 (2015).

422. Song, J. *et al.* RS-1 enhances CRISPR/Cas9- and TALEN-mediated knock-in efficiency. *Nat. Commun.* **7**, 1–7 (2016).

423. Canny, M. D. *et al.* Inhibition of 53BP1 favors homology-dependent DNA repair and increases CRISPR-Cas9 genome-editing efficiency. *Nat. Biotechnol.* **36**, 95–102 (2018).

424. Richards, D. Y. *et al.* AAV-mediated CRISPR/Cas9 gene editing in murine phenylketonuria. *Mol. Ther. - Methods Clin. Dev.* **17**, 234–245 (2020).

425. Guo, Q. *et al.* 'Cold shock' increases the frequency of homology directed repair gene editing in induced pluripotent stem cells. *Sci. Rep.* **8**, 1–11 (2018).

426. Suzuki, K. *et al.* In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature* **540**, 144–149 (2016).

427. Anguela, X. M. *et al.* Robust ZFN-mediated genome editing in adult hemophilic mice. *Blood* **122**, 3283–3287 (2013).

428. He, X. *et al.* Knock-in of large reporter genes in human cells via CRISPR/Cas9-induced homology-dependent and independent DNA repair. *Nucleic Acids Res.* **44**, (2016).

429. Jacinto, F. V., Link, W. & Ferreira, B. I. CRISPR/Cas9-mediated genome editing: From basic research to translational medicine. *J. Cell. Mol. Med.* **24**, 3766–3778 (2020).

430. Song, F. & Stieger, K. Optimizing the DNA donor template for homology-directed repair of double-strand breaks. *Mol. Ther. - Nucleic Acids* **7**, 53–60 (2017).

431. Irion, U., Krauss, J. & Nüsslein-Volhard, C. Precise and efficient genome editing in zebrafish using the CRISPR/Cas9 system. *Development* **141**, 4827–4830 (2014).

432. Li, H. *et al.* In vivo genome editing restores hemostasis in a mouse model of hemophilia. *Nature* **475**, 217–221 (2011).

433. Schwank, G. *et al.* Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. *Cell Stem Cell* **13**, 653–658 (2013).

434. Truong, D. J. J. *et al.* Development of an intein-mediated split-Cas9 system for gene therapy. *Nucleic Acids Res.* **43**, 6450–6458 (2015).

435. Yang, Y. *et al.* A dual AAV system enables the Cas9-mediated correction of a metabolic liver disease in newborn mice. *Nat. Biotechnol.* **34**, 334–338 (2016).

436. Yin, H. *et al.* Therapeutic genome editing by combined viral and non-viral delivery of CRISPR system components in vivo. *Nat. Biotechnol.* **34**, 328–333 (2016).

437. Song, C. Q. *et al.* In vivo genome editing partially restores alpha1-antitrypsin in a murine model of AAT deficiency. *Hum. Gene Ther.* **29**, 853–860 (2018).

438. Roth, T. L. *et al.* Reprogramming human T cell function and specificity with non-viral genome targeting. *Nature* **559**, 405–409 (2018).

439. Yin, H. *et al.* Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. *Nat. Biotechnol.* **32**, 551–553 (2014).

440. Lee, J. S., Kallehauge, T. B., Pedersen, L. E. & Kildegaard, H. F. Site-specific integration in CHO cells mediated by CRISPR/Cas9 and homology-directed DNA repair pathway. *Sci. Rep.* **5**, 1–11 (2015).

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

442. DeWitt, M. A. *et al.* Selection-free genome editing of the sickle mutation in human adult hematopoietic stem/progenitor cells. *Sci. Transl. Med.* **8**, 139–148 (2016).

443. Liang, X., Potter, J., Kumar, S., Ravinder, N. & Chesnut, J. D. Enhanced CRISPR/Cas9-

mediated precise genome editing by improved design and delivery of gRNA, Cas9 nuclease, and donor DNA. *J. Biotechnol.* **241**, 136–146 (2017).

444. Miura, H., Quadros, R. M., Gurumurthy, C. B. & Ohtsuka, M. Easi-CRISPR protocol for creating knock-in and conditional knockout mouse models using long ssDNA donors. *Nat Protoc* **13**, 195–215 (2018).

445. Nishimasu, H. *et al.* Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science.* **361**, 1259–1262 (2018).

446. Hu, J. H. *et al.* Evolved Cas9 variants with broad PAM compatibility and high DNA specificity Johnny. *Nature* **556**, 57–63 (2018).

447. Gao, L. *et al.* Engineered Cpf1 variants with altered PAM specificities increase genome targeting range. *Nat. Biotechnol.* **35**, 789–792 (2017).

448. Barbalat, R., Ewald, S. E., Mouchess, M. L. & Barton, G. M. Nucleic acid recognition by the innate immune system. *Annu. Rev. Immunol.* **29**, 185–214 (2011).

449. Wienert, B., Shin, J., Zelin, E., Pestal, K. & Corn, J. E. In vitro–transcribed guide RNAs trigger an innate immune response via the RIG-I pathway. *PLoS Biol.* **16**, 1–18 (2018).

450. Kim, S. *et al.* CRISPR RNAs trigger innate immune responses in human cells. *Genome Res.* **28**, 367–373 (2018).

451. Chew, W. L. *et al.* A multi-functional AAV-CRISPR-Cas9 and its host response. *Nat. Methods* **13**, 868–874 (2016).

452. Charlesworth, C. T. *et al.* Identification of preexisting adaptive immunity to Cas9 proteins in humans. *Nat. Med.* **25**, 249–254 (2019).

453. Wang, D. *et al.* Adenovirus-mediated somatic genome editing of Pten by CRISPR/Cas9 in mouse liver in spite of Cas9-specific immune responses. *Hum. Gene Ther.* **26**, 432–442 (2015).

454. Li, L., Hu, S. & Chen, X. Non-viral delivery systems for CRISPR/Cas9-based genome editing: Challenges and opportunities. *Biomaterials* **171**, 207–218 (2018).

455. Qiurong, D. *et al.* Permanent alteration of PCSK9 with in vivo CRISPR-Cas9 genome editing. *Circ. Res.* **115**, 488–492 (2014).

456. Guan, Y. *et al.* CRISPR /Cas9-mediated somatic correction of a novel coagulator factor IX gene mutation ameliorates hemophilia in mouse . *EMBO Mol. Med.* **8**, 477–488 (2016).

457. Ortinski, P. I., O'Donovan, B., Dong, X. & Kantor, B. Integrase-deficient lentiviral vector as an all-in-one platform for highly efficient CRISPR/Cas9-mediated gene editing. *Mol. Ther. - Methods Clin. Dev.* **5**, 153–164 (2017).

458. Swiech, L. *et al.* In vivo interrogation of gene function in the mammalian brain using CRISPR-Cas9. *Nat. Biotechnol.* **33**, 102–106 (2015).

459. Long, C. *et al.* Postnatal genome editing partially restores dystrophin expression in a mouse model of muscular dystrophy. *Science.* **351**, 400–403 (2016).

460. Nelson, C. E. *et al.* Long-term evaluation of AAV-CRISPR genome editing for Duchenne Muscular Dystrophy. *Nat. Med.* **176**, 139–148 (2019).

461. Kim, E. *et al.* In vivo genome editing with a small Cas9 orthologue derived from Campylobacter jejuni. *Nat. Commun.* **8**, 1–12 (2017).

462. Ran, F. A. *et al.* In vivo genome editing using Staphylococcus aureus Cas9. *Nature* **520**, 186–191 (2015).

463. Nelson, C. E. *et al.* In vivo editing improves muscle function in mouse of DMD. *Science.* **351**, 403–407 (2016).

464. Maeder, M. L. *et al.* Development of a gene-editing approach to restore vision loss in Leber congenital amaurosis type 10. *Nat. Med.* **25**, 229–233 (2019).

465. Ibraheim, R. *et al.* All-in-one adeno-associated virus delivery and genome editing by Neisseria meningitidis Cas9 in vivo. *Genome Biol.* **19**, 1–11 (2018).

466. Zhen, S. *et al.* Harnessing the clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated Cas9 system to disrupt the hepatitis B virus. *Gene Ther.* **22**, 404–412 (2015).

467. Bakondi, B. *et al.* In vivo CRISPR/Cas9 gene editing corrects retinal dystrophy in the S334ter-3 rat model of autosomal dominant retinitis pigmentosa. *Mol. Ther.* **24**, 556–563 (2016).

468. Latella, M. C. *et al.* In vivo editing of the human mutant rhodopsin gene by electroporation of plasmid-based CRISPR/Cas9 in the mouse retina. *Mol. Ther. - Nucleic Acids* **5**, e389 (2016).

469. Luo, Y. L. *et al.* Macrophage-specific in vivo gene editing using cationic lipid-assisted polymeric nanoparticles. *ACS Nano* **12**, 994–1005 (2018).

470. Wang, H. X. *et al.* Nonviral gene editing via CRISPR/Cas9 delivery by membrane-disruptive and endosomolytic helical polypeptide. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 4903–4908 (2018).

471. Zhang, Z. *et al.* Cationic polymer-mediated CRISPR/Cas9 plasmid delivery for genome editing. *Macromol. Rapid Commun.* **40**, 1–8 (2019).

472. Wang, P. *et al.* Thermo-triggered release of CRISPR-Cas9 system by lipid-encapsulated gold nanoparticles for tumor therapy. *Angew. Chemie - Int. Ed.* **57**, 1491–1496 (2018).

473. Lin, Y. *et al.* Exosome–liposome hybrid nanoparticles deliver CRISPR/Cas9 system in MSCs. *Adv. Sci.* **5**, 1–9 (2018).

474. Xu, C. F. *et al.* Rational designs of in vivo CRISPR-Cas delivery systems. *Adv. Drug Deliv. Rev.* **21**, (2019).

475. Anderson, B. R. *et al.* Nucleoside modifications in RNA limit activation of 2′-5′-oligoadenylate synthetase and increase resistance to cleavage by RNase L. *Nucleic Acids Res.* **39**, 9329–9338 (2011).

476. Vaidyanathan, S. *et al.* Uridine depletion and chemical modification increase Cas9 mRNA activity and reduce immunogenicity without HPLC purification. *Mol. Ther. - Nucleic Acids* **12**, 530–542 (2018).

477. Hendel, A. *et al.* Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. *Nat. Biotechnol.* **33**, 985–989 (2015).

478. Yin, H. *et al.* Structure-guided chemical modification of guide RNA enables potent non-viral in vivo genome editing. *Nat. Biotechnol.* **35**, 1179–1187 (2017).

479. Miller, J. B. *et al.* Non-viral CRISPR/Cas gene editing in vitro and in vivo enabled by synthetic nanoparticle co-delivery of Cas9 mRNA and sgRNA. *Angew Chem Int Ed Engl* **56**, 1059–1063 (2017).

480. Jiang, C. *et al.* A non-viral CRISPR/Cas9 delivery system for therapeutically targeting HBV DNA and pcsk9 in vivo. *Cell Res.* **27**, 440–443 (2017).

481. Zuris, J. A. *et al.* Efficient delivery of genome-editing proteins in vitro and in vivo. *Nat. Biotechnol.* **33**, 73–80 (2015).

482. Sun, W. *et al.* Efficient delivery of CRISPR-Cas9 for genome editing via self-assembled DNA nanoclews. *Angew Chem Int Ed Engl* **54**, 12029–12033 (2015).

483. Chen, Z. *et al.* Targeted delivery of CRISPR/Cas9-mediated cancer gene therapy via liposome-templated hydrogel nanoparticles. *Adv Funct Mater* **27**, (2017).

484. Lee, K. *et al.* Nanoparticle delivery of Cas9 ribonucleoprotein and donor DNA in vivo induces homology-directed DNA repair. *Nat Biomed Eng* **1**, 889–901 (2017).

485. Lee, B. *et al.* Nanoparticle delivery of CRISPR into the brain rescues a mouse model of fragile X syndrome from exaggerated repetitive behaviours. *Nat Biomed Eng* **2**, 497–507 (2018).

486. Gao, X. *et al.* Treatment of autosomal dominant hearing loss by in vivo delivery of genome editing agents. *Nature* **553**, 217–221 (2018).

487. Kim, S. M. *et al.* Simple in vivo gene editing via direct self-assembly of Cas9 ribonucleoprotein complexes for cancer treatment. *ACS Nano* **12**, 7750–7760 (2018).

488. de Buhr, H. & Lebbink, R. J. Harnessing CRISPR to combat human viral infections. *Curr. Opin. Immunol.* **54**, 123–129 (2018).

489. Pavel-Dinu, M. *et al.* Gene correction for SCID-X1 in long-term hematopoietic stem cells. *Nat. Commun.* **10**, (2019).

490. Yoshiba, T. *et al.* CRISPR/Cas9-mediated cervical cancer treatment targeting human papillomavirus E6. *Oncol. Lett.* **17**, 2197–2206 (2019).

491. Lee, H., Yoon, D. E. & Kim, K. Genome editing methods in animal models. *Animal Cells Syst. (Seoul).* **24**, 8–16 (2020).

492. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-γuided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).

493. Gilbert, L. A. *et al.* XCRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442 (2013).

494. Kim, D. *et al.* Genome-wide target specificities of CRISPR RNA-guided programmable deaminases. *Nat. Biotechnol.* **35**, 475–480 (2017).

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

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

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

498. Wu, X., Mao, S., Ying, Y., Krueger, C. J. & Chen, A. K. Progress and challenges for live-cell imaging of genomic loci using CRISPR-based platforms. *Genomics, Proteomics Bioinforma.* **17**, 119–128 (2019).

499. Chen, B., Guan, J. & Huang, B. Imaging specific genomic DNA in living cells. *Annu. Rev. Biophys.* **45**, 1–23 (2016).

500. Abudayyeh, O. O. *et al.* RNA targeting with CRISPR-Cas13a. *Nature* **550**, 280–284 (2017).

501. Shi, B. *et al.* Biodistribution of small interfering RNA at the organ and cellular levels after lipid nanoparticle-mediated delivery. *J. Histochem. Cytochem.* **59**, 727–740 (2011).

502. Tabernero, J. *et al.* First-in-humans trial of an RNA interference therapeutic targeting VEGF and KSP in cancer patients with liver involvement. *Cancer Discov.* **3**, 406–417 (2013).

503. Miao, L. *et al.* Synergistic lipid compositions for albumin receptor mediated delivery of mRNA to the liver. *Nat. Commun.* **11**, (2020).

504. Fujisaka, S. *et al.* Diet, genetics, and the gut microbiome drive dynamic changes in plasma metabolites. *Cell Rep.* **22**, 3072–3086 (2018).

505. Ren, Y. *et al.* Chronic stress disturbs metabolome of blood plasma and urine in diabetic rats. *Front. Psychiatry* **9**, 1–12 (2018).

506. Witzigmann, D. *et al.* Lipid nanoparticle technology for therapeutic gene regulation in the liver. *Adv. Drug Deliv. Rev.* (2020) doi:10.1016/j.addr.2020.06.026.

507. Aronsohn, A. I. & Hughes, J. A. Nuclear localization signal peptides enhance cationic liposome-mediated gene therapy. *J. Drug Target.* **5**, 163–169 (1998).

508. Akita, H. *et al.* Improving in vivo hepatic transfection activity by controlling intracellular trafficking: The function of GALA and maltotriose. *Mol. Pharm.* **8**, 1436–1442 (2011).

509. Khalil, I. A., Kimura, S., Sato, Y. & Harashima, H. Synergism between a cell penetrating peptide and a pH-sensitive cationic lipid in efficient gene delivery based on double-coated nanoparticles. *J. Control. Release* **275**, 107–116 (2018).

510. Bilzer, M., Roggel, F. & Gerbes, A. L. Role of Kupffer cells in host defense and liver disease. *Liver Int.* **26**, 1175–1186 (2006).

511. Poisson, J. *et al.* Liver sinusoidal endothelial cells: Physiology and role in liver diseases. *J. Hepatol.* **66**, 212–227 (2017).

512. Gebhardt, R. & Matz-Soja, M. Liver zonation: Novel aspects of its regulation and its impact on homeostasis. *World J. Gastroenterol.* **20**, 8491–8504 (2014).

513. Kietzmann, T. Metabolic zonation of the liver: The oxygen gradient revisited. *Redox Biol.* **11**, 622–630 (2017).

514. Sago, C. D., Krupczak, B. R., Lokugamage, M. P., Gan, Z. & Dahlman, J. E. Cell subtypes within the liver microenvironment differentially interact with lipid nanoparticles. *Cell. Mol. Bioeng.* **12**, 389–397 (2019).

515. Racine-Samson, L. *et al.* The metabolic organization of the adult human liver: A comparative study of normal, fibrotic, and cirrhotic liver tissue. *Hepatology* **24**, 104–113 (1996).

516. Peter Bell *et al.* Inverse zonation of hepatocyte transduction with AAV vectors between mice and non-human primates. *Mol. Genet. Metab.* **104**, 395–403 (2011).

517. Krumsiek, J. *et al.* Gender-specific pathway differences in the human serum metabolome. *Metabolomics* **11**, 1815–1833 (2015).

518. Rist, M. J. *et al.* Metabolite patterns predicting sex and age in participants of the Karlsruhe Metabolomics and Nutrition (KarMeN) study. *PLoS One* **12**, 1–21 (2017).

519. Hu, J. M. *et al.* Functional analyses of albumin expression in a series of hepatocyte cell lines and in primary hepatocytes. *Cell Growth Differ.* **3**, 577–588 (1992).

520. Matsukawa, K., Moriyama, A., Kawai, Y., Asai, K. & Kato, T. Tissue distribution of human gliostatin/platelet-derived endothelial cell growth factor (PD-ECGF) and its drug-induced expression. *Biochim. Biophys. Acta - Mol. Cell Res.* **1314**, 71–82 (1996).

521. Pauly, J. L. *et al.* Identification and comparative analysis of thymidine phosphorylase in the plasma of healthy subjects and cancer patients: Brief communication. *J. Natl. Cancer Inst.* **58**, 1587–1590 (1977).

522. Waltenberger, J., Usuki, K., Fellström, B., Funa, K. & Heldin, C. H. Platelet-derived endothelial cell growth factor Pharmacokinetics, organ distribution and degradation after intravenous administration in rats. *FEBS Lett.* **313**, 129–132 (1992).

523. Barzel, A. *et al.* Promoterless gene targeting without nucleases ameliorates haemophilia B in mice. *Nature* **517**, 360–364 (2015).

524. Vandendriessche, T. & Chuah, M. K. Hitting the target without pulling the trigger. *Mol. Ther.* **23**, 4–6 (2015).

525. Porro, F. *et al.* Promoterless gene targeting without nucleases rescues lethality of a Crigler-Najjar syndrome mouse model. *EMBO Mol. Med.* **9**, 1346–1355 (2017).

9. ANNEX

Application	Name	Orientation	Sequence (5'-3')
gRNA *in vitro* transcription	**ALB1-IVT-F**	F	TAATACGACTCACTATAGTTCCTGTAACGATCGG
	ALB1-IVT-R	R	TTCTAGCTCTAAAACGTTCCCGATCGTTACAGGA
	ALB2-IVT-F	F	TAATACGACTCACTATAGTTAGTATAGCATGGTC
	ALB2-IVT-R	R	TTCTAGCTCTAAAACGCTCGACCATGCTATACTA
	ALB3-IVT-F	F	TAATACGACTCACTATAGTGCCAGTTCCCGATCG
	ALB3-IVT-R	R	TTCTAGCTCTAAAACGTAACGATCGGGAACTGGC
	TYMP1-IVT-F	F	TAATACGACTCACTATAGGAATAGGTACTAGACG
	TYMP1-IVT-R	R	TTCTAGCTCTAAAACCCCACGTCTAGTACCTATT
	TYMP2-IVT-F	F	TAATACGACTCACTATAGAATTGAATAGGTACTA
	TYMP2-IVT-R	R	TTCTAGCTCTAAAACCGTCTAGTACCTATTCAAT
	TYMP3-IVT-F	F	TAATACGACTCACTATAGAATAGGTACTAGACGT
	TYMP3-IVT-R	R	TTCTAGCTCTAAAACGCCCACGTCTAGTACCTAT
	TYMP4-IVT-F	F	TAATACGACTCACTATAGAGCTGCCCCATACGCA
	TYMP4-IVT-R	R	TTCTAGCTCTAAAACGAGGTGCGTATGGGGCAGC
	TYMP5-IVT-F	F	TAATACGACTCACTATAGTGGATCCCTGAGGTGC
	TYMP5-IVT-R	R	TTCTAGCTCTAAAACATACGCACCTCAGGGATCC
DNA templates sequencing	**Seq-1**	F	CTCGCGCGTTTCGGTGATGA
	Seq-2	F	CGCCGAGAAGCACGTAAGAG
	Seq-3	R	AAGATGCCAGTTCCCGATCG
	Seq-4	R	AGTGAGCGCAACGCAATTAA
	Seq-5	F	AATAGGCGTATCACGAGGCC
	Seq-6	F	TCCGGAGCGCTGTAACCCGA
	Seq-7	R	GCTGCCATTGCTCCAGTCTG
	Seq-8	R	TTAATGCAGCTGGCACGACA
	Seq-9	F	GCTGCCCGACAACCACTA
	Seq-10	R	TCTCACACACAGTTCGCCTC
	Seq-11	F	GTCCTGTCGGGTTTCGCCAC
	Seq-12	R	GACCAGCTCCACGGTGCCATC
	Seq-13	R	TCAACAATGCCGTCTGCCGGTGC
	Seq-14	R	ATTGTCTCATGAGCGGATAC

Table S1. List of primers used for PCRs and sequencing.

Element	Sequence (5'-3')
***eGFP* cDNA**	ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGA CGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGAC CCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACC TACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCC ATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGC GCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAA GGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCA TGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGC AGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCC GACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATG GTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAA
h*TYMP* cDNA	ATGGCAGCCTTGATGACCCCGGGAACCGGGGCCCCACCCGCGCCTGGTGACTTCTCCGGGGAAGGG AGCCAGGGACTTCCCGACCCTTCGCCAGAGCCCAAGCAGCTCCCGGAGCTGATCCGCATGAAGCGA GACGGAGGCCGCCTGAGCGAAGCGGACATCAGGGGCTTCGTGGCCGCTGTGGTGAATGGGAGCGC GCAGGGCGCACAGATCGGGGCCATGCTGATGGCCATCCGACTTCGGGGCATGGATCTGGAGGAGA CCTCGGTGCTGACCCAGGCCCTGGCTCAGTCGGGACAGCAGCTGGAGTGGCCAGAGGCCTGGCGCC AGCAGCTTGTGGACAAGCATTCCACAGGGGGTGTGGGTGACAAGGTCAGCCTGGTCCTCGCACCT GCCCTGGCGGCATGTGGCTGCAAGGTGCCAATGATCAGCGGACGTGGTCTGGGGCACACAGGAGG CACCTTGGATAAGCTGGAGTCTATTCCTGGATTCAATGTCATCCAGAGCCCAGAGCAGATGCAAG TGCTGCTGGACCAGGCGGGCTGCTGTATCGTGGGTCAGAGTGAGCAGCTGGTTCCTGCGGACGGA ATCCTATATGCAGCCAGAGATGTGACAGCCACCGTGGACAGCCTGCCACTCATCACAGCCTCCATT CTCAGTAAGAAACTCGTGGAGGGGCTGTCCGCTCTGGTGGTGGACGTTAAGTTCGGAGGGGCCGC CGTCTTCCCCAACCAGGAGCAGGCCCGGGAGCTGGCAAAGACGCTGGTTGGCGTGGGAGCCAGCCT AGGGCTTCGGGTCGCGGCAGCGCTGACCGCCATGGACAAGCCCCTGGGTCGCTGCGTGGGCCACGC CCTGGAGGTGGAGGAGGCGCTGCTCTGCATGGACGGCGCAGGCCCGCCAGACTTAAGGGACCTGGT CACCACGCTCGGGGGCGCCCTGCTCTGGCTCAGCGGACACGCGGGGACTCAGGCCCAGGGCGCTGC CCGGGTGGCCGCGGCGCTGGACGACGGCTCGGCCCTTGGCCGCTTCGAGCGGATGCTGGCGGCGCA GGGCGTGGATCCCGGTCTGGCCCGAGCCCTGTGCTCGGGAAGTCCCGCAGAACGCCGGCAGCTGCT GCCTCGCGCCCGGGAGCAGGAGGAGCTGCTGGCGCCCGCAGATGGCACCGTGGAGCTGGTCCGGGC GCTGCCGCTGGCGCTGGTGCTGCACGAGCTCGGGGCCGGGCGCAGCCGCGCTGGGGAGCCGCTCCG CCTGGGGGTGGGCGCAGAGCTGCTGGTCGACGTGGGTCAGAGGCTGCGCCGTGGGACCCCCTGGCT CCGCGTGCACCGGGACGGCCCCGCGCTCAGCGGCCCGCAGAGCCGCGCCCTGCAGGAGGCGCTCGT ACTCTCCGACCGCGCGCCATTCGCCGCCCCCTTGCCCTTCGCAGAGCTCGTTCTGCCGCCGCAGCAA TAA
Partial m*Tymp* cDNA (exons 3-11)	GGGCCATGCTAATGGCCATTCGGCTGCAGGGAATGAACCTAGAAGAGACATCTGTGCTGACTCGG GCCCTTGCAGAGTCTGGGCAGCAACTGGAGTGGCCCAAAGCCTGGCACCAGCAGCTTGTTGACAAA CACTCCACAGGAGGTGTGGGTGACAAGGTTAGCCTGGTTCTGGCACCTGCCCTGGCTGCATGCGGC TGTAAGGTGCCAATGATCAGCGGCCGCAGTCTGGGACACACAGGGGGCACACTGGATAAGCTGGA ATCGATTCCTGGGTTCGGTGTCACCCAGAGCCCAGAGCAGATGCTGCACATCCTTGAGGAAGTCGG CTGCTGCATTGTCGGCCAGAGTGCAAAGCTGGTTCCTGCTGATGGGATCCTGTATGCTGCACGAGA TGTGACAGCCACTGTGGACAGTGTCCCACTCATTACAGCTTCAATCCTCAGTAAGAAGGCCGTGGA GGGACTGTCCACTCTGGTGGTAGATGTTAAGTTTGGGGGAGCTGCAGTCTTCCCAGACCAGGAGA AGGCCCGAGAACTGGCAAAGATGTTGGTTAGAGTGGGAGTGAGCCTGGGGCTGAAGGTCGCCGCA GCCCTGACCGCCATGGATAACCCTCTGGGCCGCAGCGTGGGCCATACCCTGGAGGTGGAAGAAGCG TTGCTTTGCCTGGATGGTGCGGGGCCGCCGGATCTGCGGGACCTGGTCATTAGGCTAGGAGGCGCC ATTCTTTGGATTAGCGGACAGGCGGAAACACAAGACCAAGGCGCCGCCCGAGTGGCCGCGGCGTT GGATGACGGCTCCGCGCGGCGCCGCTTCCAGCTGATGCTTTCGGCGCAGGGCGTGGACCCCGGCCT AGCTAAAGCATTGTGCTCCGGGAGCCCCACGCAACGCCGGCAGCTGCTGCCTCACGCCCGAGAGCA AGAGGAGCTGCTCGCACCGGCAGACGGCATTGTTGAGTGCGTCCGAGCGCTGCCGCTGGCCCGTGT GCTGCACGACCTCGGAGCCGGACGCAGCCGAGCGGGGCAGCCGATCAGACCGGGAGTGGGTGCCGA GGTACTGGTCGACGTGGGCCAGTGTTTGTCCCGAGGGACACCCTGGCTCCGCGTGCACCTAGACGG CCCTGCGCTCAGCAGTCAGCAGCGACGCACTCTCCAGGGGGCGCTCGTGCTATCAGATCGAGCACC CTTCAAGGTCCCTTCACCCTTCGCTGAACTTGTCCTCCCACCGACCATCGCACAGCCC
SA	TTCTCTCCCTGTTTCCACAG
pA (SV40)	ACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAA GCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGT

Right HA form *Tymp* (1st generation)	GTGCGGGGTGGGGGGAGTGATTTGTGAGGTCGAGAAATGGAAGCCAGATGGGGTTGGGAATGGG ACAGTGCTGGATAGAGAAAAACATGAGACTGGTGAGTGCCAGAGACCCTGTAAGAGCAAGGTTTG GACAAGTGACCATGGATTCAAAGACCTGGATGGGGTGGAGGTGGGGAGTTCCAGTGGTGTCTCAA TATCCCTTACAGACTGGAGCAATGGCAGCTCCAGGGACACCACCTCCCTCGGCCTCAGGGGGAGGA GGTGGGGAGCCCAGGCAGCTGCCTGAGCTGATCCGCCTGAAGCGAGATGGAGGGCATCTGAGAGA AGCAGATATCAGAAACTTCGTGCACGCGGTGATAGATGGAAGAGCACAGGACACACAAATTGGTG TGTGGTAACTAGACCCCAACACCCCTTCCAGCTGCCCCATACGCACCTCAGGGATCCACAGTACCC CACCACCATCACGGCCCCAGAAAACTCAGACTGCGTGAACACCTCGCTCATTGTATCCATACTTAG GGGCCATGCTAATGGCCATTCGGCTGCAGGGAATGAACCTAGAAGAGACATCTGTGCTGACTCGG GCCCTTGCAGAGTCTGGGCAGCAACTGGAGTGGCCCAAAGCCTGGCACCAGCAGCTTGTTGACAAA CACTCCACAGGAGGTGTGGGTGACAAGGTTAGCCTGGTTCTGGCACCTGCCCTGGCTGCATGCGGC TGTAAGGTTAGTGACCACGTCCTCTACAGACCTCTGACCCGCTCACAGGAACCTCCAACCCTCCAG CCTGGCTCAAAGCTCTCAAAGACAAAATCCAGGGCAGTCCCATCGCTCAATACACAAATTCCACCA TGGAGTAGA
Left HA form *Tymp* (1st generation)	TCCTATGATCTACAAGACTCGAGCCCCGCAGTTCACGATGCTGGGGAGGACTCTGCCACCCCGAGA GAACACCAAGAAGCCAGGACCTGCATCCTACAGCGTGGACAAGGTGGTCTGGAGTCGAGGGTCAA GGGGCAGGGGATGAGGGACTGAGGTCTAAGGCATGGGTCAAAGTCCCCAGGGAAATGGCATCCCT GTATGAAGCCGAACAGCTGGCTCGGTAAGGTCAGCCTCGTCTCCCCGACCCGCAGCACCGGAAGCC TTGTGGCTGGAGCTTCGGGATCAGACACTCAGACTACCTAGCCCCCATGATACAGTACCACGTGGA CGACTGATGCCTGCAAGCACGCGCGCACAGGCTCTGGGTTGTTTACTTAAAGCTTTCTAAGTCAGT GCTGTGTGTCTCTCTTTCTTTCTCCGAAGCGGGCAAGGAGGGGTGGGGGAGGAGAGGACGAGGTG AGTTCGAGCATTGATTGTCTCAGTGGTATCTGAACCTAGAGATCTGTCGCTTGCCTAGAGATCTG TCGCACAGGACAAGCAGTGCTCGGGACCCATAGATACCAAGACCCCAGGGAACTGGGCAGAGCGA ACTCCGGTCCGGAGCGCTGTAACCCGACCCTTAAAGGTCCTGGCAATGCCAATTAGCCAGATAGCA GGTGGGGAAGAGCTTAAGTGGTGATTGGCCTTACCTTGACAAGTGCCGCCTACTGGGACAGCCCT ATTTTTCCACTATTCCTAACCGTAATCCTGGACGGTGAACTGTCTCCTCTGTACCTAAGTCCACGC AGGACTGAGGGATAACAGGCTGGAACCGAGCCAGATCCAGAAAACTGGTAAAGGCCCCTGAATTG AATAGGTAC
Right HA form *Alb* (1st generation)	CTATACTAAAAATTAAAAGTGTGTGTTACTAATTTTATAAATGGAGTTTCCATTTATATTTACCT TTATTTCTTATTTACCATTGTCTTAGTAGATATTTACAAACATGACAGAAACACTAAATCTTGAG TTTGAATGCACAGATATAAACACTTAACGGGTTTTAAAAATAATAATGTTGGTGAAAAAATATA ACTTTGAGTGTAGCAGAGAGGAACCATTGCCACCTTCAGATTTTCCTGTAACGATCGGGAACTGG CATCTTCAGGGAGTAGCTTAGGTCAGTGAAGAGAAGAACAAAAAGCAGCATATTACAGTTAGTTG TCTTCATCAATCTTTAAATATGTTGTGTGGTTTTTCTCTCCCTGTTTCCACAGACAAGAGTGAGAT CGCCCATCGGTATAATGATTTGGGAGAACAACATTTCAAAGGCCTGTAAGTTATAATGCTGAAAG CCCACTTAATATTTCTGGTAGTATTAGTTAAAGTTTTAAAACACCTTTTTCCACCTTGAGTGTGA GAATTGTAGAGCAGTGCTGTCCAGTAGAAATGTGTGCATTGACAGAAAGACTGTGGATCTGTGCT GAGCAATGTGGCAGCCAGAGATCACAAGGCTATCAAGCACTTTGCACATGGCAAGTGTAACTGAG AAGCACACATTCAAATAATAGTTAATTTTAATTGAATGTATCTAGCCATGTGTGGCTAGTAGCTC CTTTCCTGGAGAGAGAATCTGGAGCCCACATCTAACTTGTTAAGTCTGGAATCTTATTTTTTATT TCTGGAAAGGTCTATGAACTATAGTTTTGGGGGCAGCTCACTTACTAACTTTTAATGCAA
Left HA form *Alb* (1st generation)	ACCCACCCCTCCCTCATACACATTCTTCATTACCTGTACATAAAAGCATATTTGGGATGAACAACC TATGCAATTCAGTTCTTAGCTTTTAAATGTGGCATGCTTCCATGCCAAGGCCCACACTGAAATGCT CAAATGGGAGACAAAGAGATTAAGCTCTTATGTAAAATTTGCTGTTTTACATAACTTTAATGAAT GGACAAAGTCTTGTGCATGGGGGTGGGGGTGGGGTTAGAGGGGAACAGCTCCAGATGGCAAACAT ACGCAAGGGATTTAGTCAAACAACTTTTTGGCAAAGATGGTATGATTTTGTAATGGGGTAGGAAC CAATGAAATGCGAGGTAAGTATGGTTAATGATCTACAGTTATTGGTTAAAGAAGTATATTAGAGC GAGTCTTTCTGCACACAGATCACCTTTCCTATCAACCCCACTAGCCTCTGGCAAAATGAAGTGGGT AACCTTTCTCCTCCTCCTCTTCGTCTCCGGCTCTGCTTTTTCCAGGGGTGTGTTTCGCCGAGAAGC ACGTAAGAGTTTTATGTTTTTTCATCTCTGCTTGTATTTTTCTAGTAATGGAAGCCTGGTATTTT AAAATAGTTAAATTTTCCTTTAGTGCTGATTTCTAGATTATTATTACTGTTGTTGTTGTTATTAT TGTCATTATTTGCATCTGAGAACCCTTAGGTGGTTATATTATTGATATATTTTTGGTATCTTTGA TGACAATAATGGGGGATTTTGAAAGCTTAGCTTTAAATTTCTTTTAATTAAAAAAAAATGCTAGG CAGAATGACTCAAATTACGTTGGATACAGTTGAATTTATTACGGTCTCATAGGGCC
Right HA form *Alb* (2nd, 3rd and 4th generation)	CTATACTAAAAATTAAAAGTGTGTGTTACTAATTTTATAAATGGAGTTTCCATTTATATTTACCT TTATTTCTTATTTACCATTGTCTTAGTAGATATTTACAAACATGACAGAAACACTAAATCTTGAG TTTGAATGCACAGATATAAACACTTAACGGGTTTTAAAAATAATAATGTTGGTGAAAAAATATA ACTTTGAGTGTAGCAGAGAGGAACCATTGCCACCTTCAGATTTTCCTGTAACGATCGGGAACTGG CATCTTCAGGGAGTAGCTTAGGTCAGTGAAGAGAAGAACAAAAAGCAGCATATTACAGTTAGTTG

	TCTTCATCAATCTTTAAATATGTTGTGTGGT
Left HA form *Alb* (2nd, 3rd and 4th generation)	GTATTTTTCTAGTAATGGAAGCCTGGTATTTTAAAATAGTTAAATTTTCCTTTAGTGCTGATTTC TAGATTATTATTACTGTTGTTGTTGTTATTATTGTCATTATTTGCATCTGAGAACCCTTAGGTGG TTATATTATTGATATATTTTTGGTATCTTTGATGACAATAATGGGGGATTTTGAAAGCTTAGCTT TAAATTTCTTTTAATTAAAAAAAAAATGCTAGGCAGAATGACTCAAATTACGTTGGATACAGTTGA ATTTATTACGGTCTCATAGGGCC

Table S2. Sequence of the elements that compose the DNA templates.

Restriction enzyme	Sequence
AgeI	ACCGGT
EcoRI	GAATTC
MluI	ACGCGT
NotI	GCGGCCGC
SpeI	ACTAGT
XbaI	TCTAGA
XhoI	CTCGAG

Table S3. List of restriction enzymes used.

Element	Sequence (5'-3')
ITR 5'	CCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGCGA CCTTTGGTCGCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTA GGGGTTCCT
ITR 3'	AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGC GACCAAAGGTCGCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCT GCCTGCAGG
EFS promoter	GGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAGGG GTCGGCAATTGATCCGGTGCCTAGAGAAGGTGGCGCGGGGTAAACTGGGAAAGTGATGTCGTGTA CTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCCGTGAACGT TCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGG
***spCas9-HF1* cDNA**	ACTAGTGTAGCCCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGG GCGTCGGGCGACCTTTGGTCGCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAA CTCCATCACTAGGGGTTCCTACGCGTGGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGCCCA CAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGATCCGGTGCCTAGAGAAGGTGGCGCGGG GTAAACTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTA TATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGGCAA GGCTGCCACCATGGATAAAAAGTATTCTATTGGTTTAGACATCGGCACTAATTCCGTTGGATGGG CTGTCATAACCGATGAATACAAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACCGT CATTCGATTAAAAAGAATCTTATCGGTGCCCTCCTATTCGATAGTGGCGAAACGGCAGAGGCGAC TCGCCTGAAACGAACCGCTCGGAGAAGGTATACACGTCGCAAGAACCGAATATGTTACTTACAAG AAATTTTTAGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTTCACCGTTTGGAAGAGTCCTTC CTTGTCGAAGAGGACAAGAAACATGAACGGCACCCCATCTTTGGAAACATAGTAGATGAGGTGGC ATATCATGAAAAGTACCCAACGATTTATCACCTCAGAAAAAAGCTAGTTGACTCAACTGATAAAG CGGACCTGAGGTTAATCTACTTGGCTCTTGCCCATATGATAAAGTTCCGTGGGCACTTTCTCATTG

	AGGGTGATCTAAATCCGGACAACTCGGATGTCGACAAACTGTTCATCCAGTTAGTACAAACCTAT AATCAGTTGTTTGAAGAGAACCCTATAAATGCAAGTGGCGTGGATGCGAAGGCTATTCTTAGCGC CCGCCTCTCTAAATCCCGACGGCTAGAAAACCTGATCGCACAATTACCCGGAGAGAAGAAAAATG GGTTGTTCGGTAACCTTATAGCGCTCTCACTAGGCCTGACACCAAATTTTAAGTCGAACTTCGACT TAGCTGAAGATGCCAAATTGCAGCTTAGTAAGGACACGTACGATGACGATCTCGACAATCTACTG GCACAAATTGGAGATCAGTATGCGGACTTATTTTTGGCTGCCAAAAACCTTAGCGATGCAATCCT CCTATCTGACATACTGAGAGTTAATACTGAGATTACCAAGGCGCCGTTATCCGCTTCAATGATCA AAAGGTACGATGAACATCACCAAGACTTGACACTTCTCAAGGCCCTAGTCCGTCAGCAACTGCCTG AGAAATATAAGGAAATATTCTTTGATCAGTCGAAAAACGGGTACGCAGGTTATATTGACGGCGGA GCGAGTCAAGAGGAATTCTACAAGTTTATCAAACCCATATTAGAGAAGATGGATGGGACGGAAGA GTTGCTTGTAAAACTCAATCGCGAAGATCTACTGCGAAAGCAGCGGACTTTCGACAACGGTAGCA TTCCACATCAAATCCACTTAGGCGAATTGCATGCTATACTTAGAAGGCAGGAGGATTTTTATCCG TTCCTCAAAGACAATCGTGAAAAGATTGAGAAAATCCTAACCTTTCGCATACCTTACTATGTGGG ACCCCTGGCCCGAGGGAACTCTCGGTTCGCATGGATGACAAGAAAGTCCGAAGAAACGATTACTC CCTGGAATTTTGAGGAAGTTGTCGATAAAGGTGCGTCAGCTCAATCGTTCATCGAGAGGATGACC GCCTTTGACAAGAATTTACCGAACGAAAAAGTATTGCCTAAGCACAGTTTACTTTACGAGTATTT CACAGTGTACAATGAACTCACGAAAGTTAAGTATGTCACTGAGGGCATGCGTAAACCCGCCTTTC TAAGCGGAGAACAGAAGAAAGCAATAGTAGATCTGTTATTCAAGACCAACCGCAAAGTGACAGTT AAGCAATTGAAAGAGGACTACTTTAAGAAAATTGAATGCTTCGATTCTGTCGAGATCTCCGGGGT AGAAGATCGATTTAATGCGTCACTTGGTACGTATCATGACCTCCTAAAGATAATTAAAGATAAGG ACTTCCTGGATAACGAAGAGAATGAAGATATCTTAGAAGATATAGTGTTGACTCTTACCCTCTTT GAAGATCGGGAAATGATTGAGGAAAGACTAAAAACATACGCTCACCTGTTCGACGATAAGGTTAT GAAACAGTTAAAGAGGCGTCGCTATACGGGCTGGGGAGCCTTGTCGCGGAAACTTATCAACGGGA TAAGAGACAAGCAAAGTGGTAAAACTATTCTCGATTTTCTAAAGAGCGACGGCTTCGCCAATAGG AACTTTATGGCCCTGATCCATGATGACTCTTTAACCTTCAAAGAGGATATACAAAAGGCACAGGT TTCCGGACAAGGGGACTCATTGCACGAACATATTGCGAATCTTGCTGGTTCGCCAGCCATCAAAA AGGGCATACTCCAGACAGTCAAAGTAGTGGATGAGCTAGTTAAGGTCATGGGACGTCACAAACCG GAAAACATTGTAATCGAGATGGCACGCGAAAATCAAACGACTCAGAAGGGGCAAAAAAACAGTCG AGAGCGGATGAAGAGAATAGAAGAGGGTATTAAAGAACTGGGCAGCCAGATCTTAAAGGAGCAT CCTGTGGAAAATACCCAATTGCAGAACGAGAAACTTTACCTCTATTACCTACAAAATGGAAGGGA CATGTATGTTGATCAGGAACTGGACATAAACCGTTTATCTGATTACGACGTCGATCACATTGTAC CCCAATCCTTTTTGAAGGACGATTCAATCGACAATAAAGTGCTTACACGCTCGGATAAGAACCGA GGGAAAAGTGACAATGTTCCAAGCGAGGAAGTCGTAAAGAAAATGAAGAACTATTGGCGGCAGCT CCTAAATGCGAAACTGATAACGCAAAGAAAGTTCGATAACTTAACTAAAGCTGAGAGGGGTGGCT TGTCTGAACTTGACAAGGCCGGATTTATTAAACGTCAGCTCGTGGAAACCCGCGCCATCACAAAG CATGTTGCGCAGATACTAGATTCCCGAATGAATACGAAATACGACGAGAACGATAAGCTGATTCG GGAAGTCAAAGTAATCACTTTAAAGTCAAAATTGGTGTCGGACTTCAGAAAGGATTTTCAATTCT ATAAAGTTAGGGAGATAAATAACTACCACCATGCGCACGACGCTTATCTTAATGCCGTCGTAGGG ACCGCACTCATTAAGAAATACCCGAAGCTAGAAAGTGAGTTTGTGTATGGTGATTACAAAGTTTA TGACGTCCGTAAGATGATCGCGAAAAGCGAACAGGAGATAGGCAAGGCTACAGCCAAATACTTCT TTTATTCTAACATTATGAATTTCTTTAAGACGGAAATCACTCTGGCAAACGGAGAGATACGCAAA CGACCTTTAATTGAAACCAATGGGGAGACAGGTGAAATCGTATGGGATAAGGGCCGGGACTTCGC GACGGTGAGAAAAGTTTTGTCCATGCCCCAAGTCAACATAGTAAAGAAAACTGAGGTGCAGACCG GAGGGTTTTCAAAGGAATCGATTCTTCCAAAAAGGAATAGTGATAAGCTCATCGCTCGTAAAAAG GACTGGGACCCGAAAAAGTACGGTGGCTTCGATAGCCCTACAGTTGCCTATTCTGTCCTAGTAGT GGCAAAAGTTGAGAAGGGAAAATCCAAGAAACTGAAGTCAGTCAAAGAATTATTGGGGATAACG ATTATGGAGCGCTCGTCTTTTGAAAAGAACCCCATCGACTTCCTTGAGGCGAAAGGTTACAAGGA AGTAAAAAAGGATCTCATAATTAAACTACCAAAGTATAGTCTGTTTGAGTTAGAAAATGGCCGAA AACGGATGTTGGCTAGCGCCGGAGAGCTTCAAAAGGGGAACGAACTCGCACTACCGTCTAAATAC GTGAATTTCCTGTATTTAGCGTCCCATTACGAGAAGTTGAAAGGTTCACCTGAAGATAACGAACA GAAGCAACTTTTTGTTGAGCAGCACAAACATTATCTCGACGAAATCATAGAGCAAATTTCGGAAT TCAGTAAGAGAGTCATCCTAGCTGATGCCAATCTGGACAAAGTATTAAGCGCATACAACAAGCAC AGGGATAAACCCATACGTGAGCAGGCGGAAAATATTATCCATTTGTTTACTCTTACCAACCTCGG CGCTCCAGCCGCATTCAAGTATTTTGACACAACGATAGATCGCAAACGATACACTTCTACCAAGG AGGTGCTAGACGCGACACTGATTCACCAATCCATCACGGGATTATATGAAACTCGGATAGATTTG TCACAGCTTGGGGGTGACGGATCCCCCAAGAAGAAGAGGAAAGTCTCGAGCGACTACAAAGACCA TGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGGCTGCAGGATGA
pA (β-globin)	AATAAAGGAAATTTATTTTCATTGCAATAGTGTGTTGGAATTTTTTGTGTCTCTCA
U6 promoter	GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAAT TGGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAA TTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAAC

	TTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACC
Tymp5 gRNA scaffold	TGGATCCCTGAGGTGCGTATGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTA TCAACTTGAAAAAGTGGCACCGAGTCGGTGC
Alb2 gRNA scaffold	AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGC GACCAAAGGTCGCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCT GCCTGCAGG
Terminator	TTTTTT

Table S4. Sequence of the elements that compose the rAAV2/8 vectors of the dual AAV approach.