The University Of Sheffield.

AAV-mediated gene editing as a therapeutic strategy for C9orf72-linked ALS/FTD

João Miguel da Conceição Alves Cruzeiro

Thesis submitted for the degree of Doctor of Philosophy (PhD)

Sheffield Institute for Translational Neuroscience

August 2020

Abstract

A hexanucleotide repeat expansion (GGGGCC or G4C2) in the first intron of Chromosome 9 open reading frame 72 (C9orf72) is the most common genetic cause for both Amyotrophic lateral sclerosis (ALS) and Frontotemporal dementia (FTD). The pathophysiology of C9orf72-ALS/FTD is thought to result from the long-term synergistic combination of C9orf72 haploinsufficiency with gain-of-function mechanisms. These are triggered by the accumulation of transcriptional and translational by-products of the repeat expansion, RNA Foci and Dipeptide repeat proteins (DPRs), respectively. Despite decades of research, ALS and FTD patients are still offered a dramatically small and inefficient list of therapeutic options that fails to significantly slow down disease progression. This highlights the urgent need for more targeted and innovative therapies that act directly on the molecular origins of these fatal and devastating neurodegenerative disorders.

In this context, we proposed to develop and test two new experimental treatments, both consisting of a highly targeted gene editing strategy using the CRISPR-Cas9 tool, to either excise the G4C2 expansion or a ~90 bp regulatory sequence located upstream of the repeats.

This thesis describes the rational design of these approaches and their practical application in proof-of-concepts using relevant cellular and animal models. In particular, we provide detailed descriptions of the design and cloning of *Staphylococcus aureus* Cas9 (SaCas9)-based expression constructs and single-guide RNAs (sgRNAs), as well as a proof-of-concept in a human cell line. Then, we report using Adeno-associated virus type 9 (AAV9) to deliver the constructs into cultured neurons, where they mediate reproducible editing of a human expanded C9orf72 allele and a decrease of RNA Foci and poly-(GP) DPRs. We also describe the testing of these viral vectors in C9-500 BAC mice, with one of our approaches mediating excision of the G4C2 expansion and reducing accumulation of poly-(GP) DPRs in the brain. In parallel, we report a severe background-associated neurological syndrome in our C9-500 BAC colony that confounds ALS/FTD behavioral assessment. Finally, we demonstrate how strategic changes to our approaches can remarkably increase their efficiency *in vitro* and *in vivo*. In conclusion, this thesis provides the basis, early development and proof-of-concepts for a new highly targeted therapy for C9orf72-ALS/FTD.

This thesis is dedicated to my mother Luisa and my father Joaquim, for their unconditional love and support throughout my life

Acknowledgements

I would first like to thank my primary supervisor, Professor Mimoun Azzouz, for making this thesis possible in the first place, for his undaunted faith in me and the project and for his relentless enthusiasm to use science for the benefit of patients. I would also like to thank my secondary supervisor Dr Laura Ferraiuolo for her continued support, valuable advice, patience and encouragement.

A special thank you to Dr Christopher Webster, Dr Evangelia Karyka, Mr Ian Coldicott, Dr James Alix, Ms Sneha Simon, Dr Adrian Higginbottom, Dr Claudia Bauer, Dr Lara Marrone, Mr Paolo Marchi, Dr Michela Pulix, Dr Lydia Castelli, Dr Guillaume Hautbergue, Ms Noemi Gatto and Ms Monika Myszczynska for having abdicated their own time to, in one way or another, actively helping with my PhD. This was either by training, performing experiments, or assisting with lengthy protocols. I would also like to extend my warm thank you to the remaining members of the Azzouz lab, including affiliates and alumni, for the constant support and the frequent laughs that made day-to-day life in the lab much more enjoyable, in particular to Dr Joseph Scarrott, Ms Louise Whiteley, Ms Emily Graves, Ms Charlotte Mason and Mr Matthew Roach. Also, a big thank you to the many people at SITraN who have gone out of their way to support and encourage me, who are too many to name. To all of these people, I am forever grateful for the vital role they played in this period of my life.

I am also forever indebted to my parents Joaquim and Luisa, my brother Luis, and my entire family and friends from Portugal, who have shaped me into the person that I am today, and without whom this would not have been possible. I have missed them every single day of my PhD. Finally, I would like to thank Emma, whose love and support in these last few years were the single most important factor to the success of this thesis. I would like to thank her from the bottom of my heart for being there for me in every single day of this PhD, the good and the bad ones, holding my hand and guiding me in the right direction every step of the way. This thesis is as much yours as it is mine.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

A	Adenine
A	Alanine
aa	Amino acids
AAV	Adeno-associated virus
AAV PHP	AAV Paul H. Patterson
AAV PHP.eB	AAV PHP enhanced B
AAV9	AAV serotype 9
AAVrh10	AAV rhesus 10
ABE	Adenine base editor
AD	Alzheimer's disease
ALS	Amyotrophic lateral sclerosis
ANOVA	Analysis of variance
ASO	Antisense oligonucleotide
ATM	Ataxia telangiectasia mutated
BAC	Bacterial Artificial Chromosome
BBB	Blood-brain barrier
BCA	Bicinchoninic acid
bGH	Bovine growth hormone
BIBD	Basophilic inclusion body disease
bp	Base pair
BSA	Bovine serum albumin
bvFTD	Behavioural-variant FTD
C	Cytosine
C57BL/6	C57 black 6
C9orf72	Chromosome 9 open reading frame 72
C9orf72-L	C9orf72 long isoform
C9orf72-S	C9orf72 short isoform
Cas9	CRISPR-associated protein 9

CBD	Corticobasal degeneration
CBE	Cytosine base editor
CBS	Corticobasal syndrome
cDNA	Complementary DNA
CHOP	Children's Hospital of Philadelphia
CIP	Calf intestine phosphatase
CjCas9	*Campylobacter jejuni* Cas9
CMAP	Compound muscle action potential
CMV	Cytomegalovirus
CNS	Central Nervous System
CRISPR	Clustered regularly interspaced short palindromic repeats
CRISPRi	CRISPR interference
crRNA	CRISPR RNA
CSF	Cerebrospinal fluid
Csy4	CRISPR-associated protein, Csy4 family
dCas9	Catalytically-dead Cas9
DCTN1	Dynactin subunit 1
DENN	Differentially expressed in normal and neoplastic cells
DEPC	Diethyl pyrocarbonate
DIV	Days *in vitro*
DMEM	Dulbecco's Modified Eagle's Medium
DNA	Deoxyribonucleic acid
DPR	Dipeptide repeat protein
dsDNA	Double-stranded DNA
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EFS1a	Elongation factor short 1a
ELISA	Enzyme-linked immunosorbent assay
EMG	Electromyography
ER	Endoplasmic reticulum
FAH	Fumarylacetoacetate hydrolase
fALS	Familial ALS
FBS	Fetal bovine serum
FDA	Food and Drug Administration

FISH	Fluorescence *in situ* Hybridization
FTD	Frontotemporal dementia
FTLD	Frontotemporal lobar degeneration
FUS	Fused in sarcoma
FVB/N	Friend Virus B NIH
FVB/NJ	Friend Virus B NIH Jackson
G	Guanine
G	Glycine
GAP	GTPase activating protein
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
gDNA	Genomic DNA
GDNF	Glial-derived neurotrophic factor
GDP	Guanosine diphosphate
GEF	GDP-GTP exchange factors
GFP	Green fluorescent protein
GPI	Glycosylphosphatidylinositol
GTP	Guanosine triphosphate
GWAS	Genome-wide association study
HA	Hemagglutinin
HBSS	Hanks' Balanced Salt Solution
HCl	Hydrochloric acid
HD	Huntington’s disease
HDR	Homology directed repair
HDV	Hepatitis delta virus
HEK	Human embryonic kidney
HGF	Hepatocyte growth factor
HH	Hammerhead
HIV	Human immunodeficiency virus
hnRNP	Heterogeneous nuclear ribonucleoprotein
HSV	Herpes simplex virus
ICD	International Classification of Diseases
ICV	Intracerebroventricular
IDT	Integrated DNA technologies
IGF-1	Insulin-like growth factor-1

Indels	Insertions/ deletions
iPSC	Induced pluripotent stem cell
ITR	Inverted terminal repeat
Kb	Kilo base
KDa	Kilodalton
KO	Knockout
LAMP1	Lysosomal associated membrane protein 1
LC3	Microtubule-associated protein 1 light chain 3
LCD	Low-complexity domains
LCL	Lymphoblastoid cell line
LLPS	Liquid-liquid phase separation
LNA	Locked nucleic acid
LY6A	Lymphocyte activation protein-6A
M	Molar
MAP2	Microtubule-associated protein 2
MAPT	Microtubule-associated protein tau
MCU	Mitochondrial Ca2+ uniporter
MeCP2	Methyl-CpG-binding protein-2
MeP	MeCP2 promoter
$MgCl_2$	Magnesium chloride
miRNA	MicroRNA
MND	Motor neuron disease
MOI	Multiplicity of infection
MRE11	Meiotic recombination 11
mRNA	Messenger RNA
MSD	Meso scale discovery
MUP	Motor unit potentials
NCBI	National Center for Biotechnology Information
NCS	Nerve conduction studies
NEB	New England Biolabs
nfvPPA	Non-fluent variant primary progressive aphasia
NGS	Next-generation sequencing
NHEJ	Non-homologous end joining
NHP	Non-human primates

NIFID	Neuronal intermediate filament inclusion disease
NIH	National institutes of health
NMJ	Neuromuscular junction
NP-40	Nonidet P-40
OCD	Obsessive-compulsive disorder
OCT	Optimal cutting temperature
OMIM	Online Mendelian Inheritance in Man
P	Proline
pAAV	Plasmid AAV
PAM	Protospacer adjacent motif
PAMmers	PAM-presenting oligonucleotides
PBP	Progressive Bulbar Palsy
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PD	Parkinson's disease
PEG	Polyethylene glycol
pegRNA	Prime editing guide RNA
PFA	Paraformaldehyde
PKR	RNA-dependent protein kinase
PLS	Primary lateral sclerosis
PMA	Progressive muscular atrophy
PPA	Primary progressive aphasias
pre-crRNA	Precursor CRISPR RNA
PSP-S	Progressive supranuclear palsy syndrome
PVDF	Polyvinylidene difluoride
qPCR	Quantitative PCR
R	Arginine
RAN	Repeat-associated non-ATG
RanGAP1	Ran-GTPase activating protein 1
RBP	RNA-binding protein
RIPA	Radioimmunoprecipitation
RNA	Ribonucleic acid
RNAse	Ribonuclease
RT	Reverse transcriptase

RTL	Ready to load
SaCas9	*Staphylococcus aureus* Cas9
sALS	Sporadic ALS
SAN	Salt-activated nuclease
SCA8	Spinocerebellar ataxia type 8
scAAV	Self-complementary AAV
SDS	Sodium dodecyl sulphate
sgRNA	Single-guide RNA
shRNA	Short-hairpin RNA
SJL/J	Swiss Jim Lambert Jackson
SMA	Spinal muscular atrophy
SMCR8	Smith-Magenis Syndrome Chromosome Region candidate 8
SNP	Single nucleotide polymorphism
SOD1	Superoxide dismutase 1
SpCas9	*Streptococcus pyogenes* Cas9
ssDNA	Single-stranded DNA
svPPA	Semantic variant primary progressive aphasia
T	Thymine
TAE	Tris-acetate EDTA
TALENS	Transcription activator-like effector nucleases
TBK1	TANK-binding kinase 1
TBST	Tris-buffered saline tween
TDP-43	Transactive response DNA-binding protein 43
TOPO	Topoisomerase
tracrRNA	Trans-activating CRISPR RNA
Tris	Tris(hydroxymethyl)aminomethane
tRNA	Transfer RNA
U	Uracil
UBQLN2	Ubiquilin-2
UCL	University College London
UGI	Uracil glycosylase inhibitor
ULK1	Unc-51-like kinase 1
Unc119	Uncoordinated-119
UPS	Ubiquitin-proteasome system

VEGF	Vascular endothelial growth factor
vg	Viral genome copies
WDR41	WD repeat-containing protein 41
WGS	Whole-genome sequencing
WT	Wild type
ZFN	Zinc-finger nucleases

1. Introduction

1.1 Amyotrophic lateral sclerosis

1.1.1 Clinical overview of ALS

Amyotrophic lateral sclerosis (ALS) is a devastating neurodegenerative condition also known as Lou Gehrig's disease and Charcot disease, respectively referring to the famous baseball player who died from ALS and to the father of French neurology who was the first to describe and name the disorder (Charcot, 1874; Rowland, 2001). ALS is a motor neuron disease (MND), and the two terms are often used interchangeably, but in the UK the term MND describes a broader group of motor neuron disorders that predominantly affect the upper motor neurons, lower motor neurons or both (Al-Chalabi et al., 2016). Indeed, MNDs include not only ALS but also Primary lateral sclerosis (PLS), Progressive muscular atrophy (PMA), Progressive Bulbar Palsy (PBP), flail arm syndrome, and flail leg syndrome, of which ALS is the most common (Al-Chalabi et al., 2016; Salameh et al., 2015). Notwithstanding this, it is not uncommon that an initial diagnosis of the abovementioned MNDs can still later evolve into an ALS diagnosis, and significant discussion still exits on whether they are separate conditions of their own or whether they all constitute incomplete forms of ALS (Al-Chalabi et al., 2016). Such confusion and overlap in the terminology stems from the highly complex clinical presentation of ALS, which can vary widely between patients. Nevertheless, the core characteristics that are overwhelmingly common in ALS patients entail dysfunction and neurodegeneration of both upper and lower motor neurons causing severe muscle weakness and atrophy, significantly limiting the movements of the patients (Figure 1-1) (Salameh et al., 2015). On rare occasions, the first symptoms can start to appear before 45 years, but the average age of onset is approximately 65 years. Being relentlessly progressive, ALS ultimately results in death, usually from respiratory failure, with only 15-20% of patients surviving beyond 5 years after the disease onset and only a small percentage living more than 10 years (Talbot, 2009).

Although ALS inevitably causes symptoms of both upper and lower motor neuron degeneration, the first regions to be affected can vary between patients, resulting in different initial presentations. Considering this, 65% of patients display what is called limb onset ALS, in which limb signs constitute the initial symptoms of the disease, whereas 30% exhibit

dysphagia and dysarthria as first signs, called bulbar onset, and 5% begin experiencing respiratory difficulties (Hardiman et al., 2011). Of these initial presentations, bulbar onset is usually associated with worse prognosis, but most affected individuals invariably end up with the combination of both upper and lower motor neuron signs that characterize this disorder: increased muscle tone/ spasticity, hyperreflexia, Babinski/ extensor plantar reflex, dysarthria, dysphagia and muscle weakness, caused by damage to the upper motor neurons; fasciculations, cramps, muscle wasting, weakness, flaccid dysarthria, resulting from damage to lower motor neurons (Kinsley and Siddique, 1993; Taylor et al., 2016). Also, contrary to the common misconception that cognition is not affected, several different studies consistently found dysfunctional behavior and cognition in a subset of ALS patients. Indeed, through language tests and detailed neuropsychological analysis, it was reported that a significant portion of these patients suffered from impaired word generation, working memory, problem-solving skills, and personality changes resembling Frontotemporal dementia (FTD) (Lomen-Hoerth et al., 2003; Portet et al., 2001; Robinson et al., 2006; Strong et al., 1999). These frequent cognitive, language and social deficits, in addition to the characteristic upper and lower motor neuron degeneration, indicate a widespread dysfunction of the entire Central Nervous System (CNS) and led to the reclassification of ALS as a multi-system neurodegenerative disorder, rather than a neuromuscular one (Crockford et al., 2018; Grossman, 2019; Hardiman et al., 2017).

The symptomatic variability described above, particularly in the early stages of the disorder, increases the complexity and delay of diagnosis, which usually takes more than one clinician and at least one year, on average, to achieve (Hardiman et al., 2011; Paganoni et al., 2014; Talbot, 2009). By lack of a definitive test, diagnosis usually relies on a set of standardized guidelines known as the El Escorial criteria. Initially published in 1994 (Brooks, 1994), El Escorial is arguably the system that has gathered the most consensus throughout the years, initially establishing four levels of increasing diagnostic certainty depending on the identification of upper and lower motor neuron signs in distinct body regions: suspected ALS, possible ALS, probable ALS and definite ALS. These guidelines were then revised twice in 2000 and 2008 with the Arlie House (Brooks et al., 2000) and Awaji-Shima criteria (de Carvalho et al., 2008), respectively, which have introduced changes to improve sensitivity and include standardized electrophysiological methods such as needle electromyography (EMG) and nerve conduction studies (NCS). In ALS patients, these methods usually reveal abnormal motor unit potentials (MUPs) due to decreased motor unit recruitment, as well as decreased amplitude of compound muscle action potential (CMAPs), detected by EMG and NCS,

respectively (de Carvalho et al., 2008; Joyce and Carter, 2013; JR, 2000). Despite being the most adopted method, El Escorial is still subject to criticism, and clinicians still rely on the International Classification of Diseases (ICD) as an alternative, as well as their own personal method, a situation which has sparked calls for a more unifying, modern and useful classification (Al-Chalabi et al., 2016).

The broad pathology of ALS is typically characterized by the atrophy of skeletal muscle and the widespread neurodegeneration found in the motor cortex, brainstem and the spinal cord, as well as diffuse astrogliosis and microgliosis. Also, similarly to other neurodegenerative diseases, post-mortem immunohistochemistry revealed the characteristic presence of protein inclusions, often ubiquitinated, in affected neuronal populations (Taylor et al., 2016). In an effort to characterize the protein content of these inclusions, Neumann and colleagues (Neumann et al., 2006) revealed that TAR DNA-binding protein 43 (TDP-43) is a key component, mislocalizing from the nucleus into the cytoplasm in diseased neurons. This pivotal work established TDP-43 proteinopathy as a major histopathological hallmark of ALS and is found in ~97% of cases, but specific ALS subtypes can display additional or different aggregation profiles (Hardiman et al., 2017). An example of this is ALS/FTD linked to Chromosome 9 open reading frame 72 (C9orf72), in which RNA Foci (DeJesus-Hernandez et al., 2011) and inclusions of dipeptide repeat proteins (DPRs) (Ash et al., 2013; Gendron et al., 2013; Mann et al., 2013; Mori et al., 2013b, 2013a; Zu et al., 2013) can be found in addition to the TDP-43 pathology. Additionally, Superoxide dismutase 1 (SOD1)- and Fused in sarcoma (FUS)-linked ALS cases display aggregated SOD1 (Kato et al., 1996; Shibata et al., 1996) and FUS (Kwiatkowski et al., 2009; Vance et al., 2009) proteins, respectively, instead of TDP-43 (Figure 1-1).

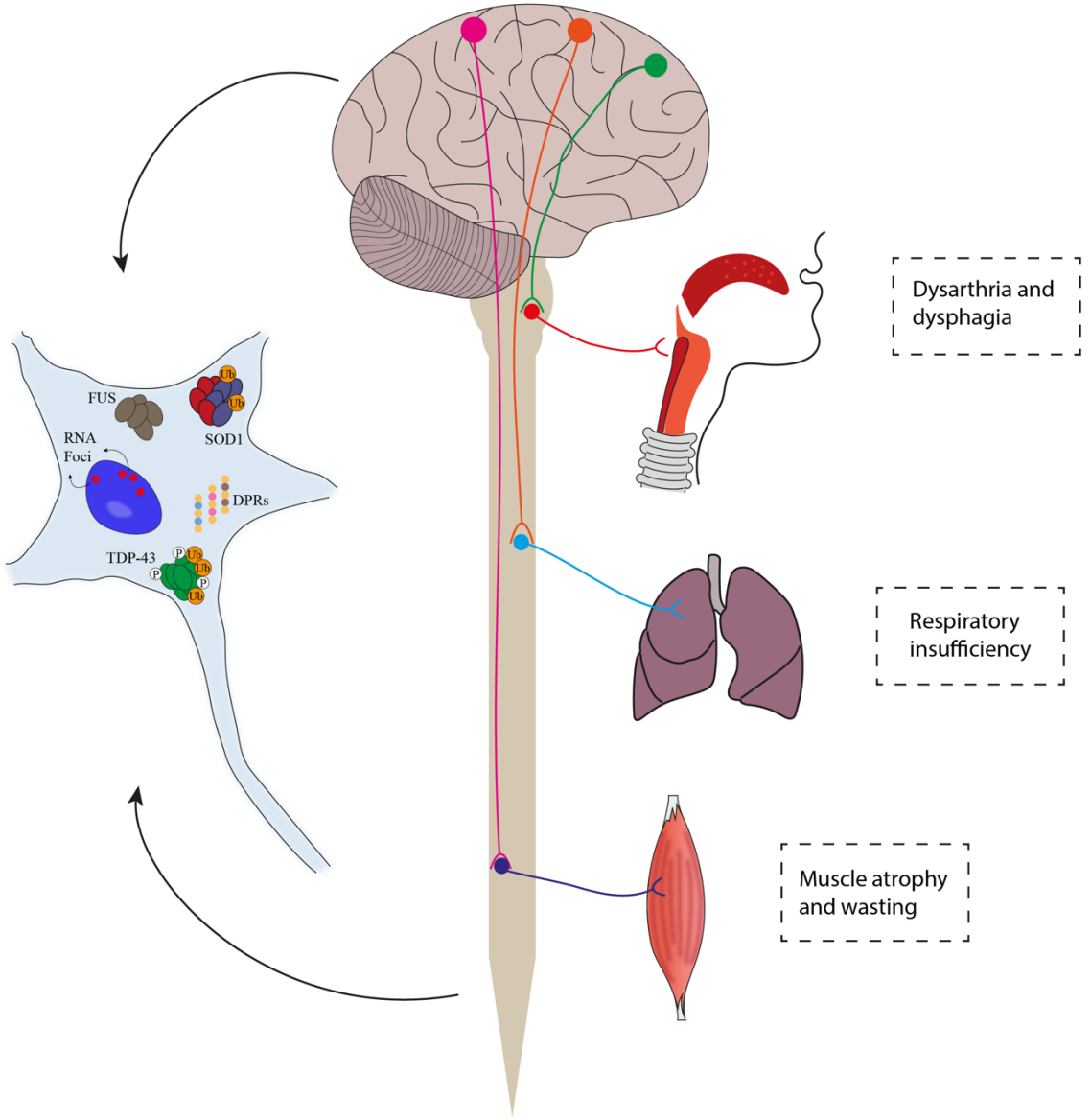


Figure 1-1. Main symptoms and neuropathology in ALS. The hallmark pathology of ALS is characterised by degeneration of upper motor neurons projecting from the motor cortex and synapsing into the brainstem and the spinal cord, as well as lower motor neurons projecting from the brainstem and spinal cord into the skeletal muscle. Degeneration of the corticospinal tract has multiple and severe consequences, of which the most critical include progressive dysfunctions in speech and swallowing (dysarthria and dysphagia), muscle wasting and paralysis and respiratory insufficiency. At the cellular level, neurons and glia accumulate inclusions of variable protein composition. In ~97% of cases, the main aggregated species is TDP-43, which is usually hyperphosphorylated and ubiquitinated, but a smaller subset of patients can instead have aggregated FUS or SOD1, usually caused by mutations in

their respective genes. Cases with mutated C9orf72 also display RNA Foci and DPR accumulation, in addition to the standard TDP-43 inclusions.

Despite the devastating socioeconomical impact of ALS, particularly on the patients and their families, there are only two disease-modifying therapies currently approved, Riluzole and Edaravone, with modest effects on disease progression and a maximum survival extension of a few months (Abe et al., 2017; Bensimon et al., 1994; Jaiswal, 2019). A pharmacological and non-pharmacological multidisciplinary approach involving specialised clinicians, therapists and caregivers can be effective in managing the symptoms of patients, reducing hospital admissions, improving the quality of life and even extending survival (Hardiman et al., 2017).

1.1.2 Epidemiology of ALS

Several studies throughout the past three decades have tried to ascertain important epidemiological aspects of ALS concerning its incidence, prevalence, age at onset and risk factors, among others, in individual regions of the globe (Alonso et al., 2009; Annegers et al., 1991; Benjaminsen et al., 2018; Bonaparte et al., 2007; Chancellor and Warlow, 1992; Forbes et al., 2007; Jun et al., 2019; Leighton et al., 2019; Logroscino et al., 2005, 2010; Longinetti et al., 2018; Mehta et al., 2018a, 2018b; Palese et al., 2019). Systematic reviews that compile and interpret these individual reports can offer an invaluable worldwide perspective, despite obvious obstacles such as different diagnosis criteria used and a lack of data from many countries (Chiò et al., 2013; Cronin et al., 2007; Longinetti and Fang, 2019). Additionally, careful analysis of the differences between these reviews can also reveal interesting global trends throughout the years. Indeed, the median incidence of ALS was estimated in 2013 to vary between 0.6 and 2.08 per 100,000 person-years in different continents, whereas the worldwide prevalence varied between 2.34 and 5.40 per 100,000 people (Chiò et al., 2013). Nevertheless, more recent data reported a worldwide incidence of 0.6 to 3.8 per 100,000 person-years and a prevalence of 4.1 to 8.4 per 100,000 people, suggesting that the number of

ALS cases is increasing (Longinetti and Fang, 2019). This is most likely due to the increase in the ageing of the world population, particularly in highly populated developing countries, a trend which is projected to continue until 2040 (Arthur et al., 2016). Data from the abovementioned studies also suggests epidemiological differences by geographical region, gender and ethnicity, with European and Asian populations being the most and the least affected ones, respectively, while caucasian males are the most affected group. Despite their undeniable importance, these conclusions should be handled with caution, as the number of existing studies varies greatly in different geographical regions and so does the methodology employed.

Studying the causes of ALS has proven to be an even more daunting endeavor, and decades of research have so far failed to determine a definitive cause for most occurrences. Approximately 90-95% of cases do not have an associated family history and are usually termed sporadic ALS (sALS), which is thought to arise from a combination of genetic and environmental risk factors that conspire with time to trigger the disease (Al-Chalabi and Hardiman, 2013; Salameh et al., 2015). Conversely, the other ~5-10% of cases have an identified genetic origin and are transmitted in a Mendelian manner, commonly referred to as familial ALS (fALS) for convenience purposes (Salameh et al., 2015). The possible environmental causes associated with sALS have been the focus of intense research throughout the years, with studies suggesting links between increased disease susceptibility and smoking (Armon, 2003, 2009; Gallo et al., 2009), exposure to pesticides (Kamel et al., 2012; Malek et al., 2012) and heavy metals (Johnson and Atchison, 2009), professional football (Chio, 2005; Chiò et al., 2009), military service (Haley, 2003; Horner et al., 2003; Kasarskis et al., 2009), and many others. Despite the described efforts, most of these associations have weak evidence to support them and have been challenged by contradicting reports and intense criticism, highlighting the extreme complexity of this undertaking (Al-Chalabi and Hardiman, 2013). However, considering its vital importance for better understanding the pathophysiological mechanisms of sALS, we share the view that stronger, unbiased population-based studies that take genetic and environmental risk factors in an integrated manner are urgently needed worldwide (Al-Chalabi and Hardiman, 2013).

1.1.3 Genetics of ALS

In addition to environmental factors, ALS has an important genetic component, which is not restricted to the 5-10% of Mendelian-inherited familial cases but also to sporadic ones. Indeed, in sALS, genetic contribution is still a key component, either through de novo mutations in critical genes or through an "at risk" genetic architecture that acts in combination with environmental factors, eventually triggering the disease (Hardiman et al., 2017). In addition to this, patients initially thought to be sporadic can have Mendelian-inherited variants that might have not been detected for several reasons including low family size, early death of family relatives for unrelated reasons, low penetrance of the mutation involved and incomplete family history, among other factors (Chiò et al., 2014).

A search in the Online Mendelian Inheritance in Man (OMIM) database reveals more than 30 genes whose mutations have been reported to either cause or significantly increase susceptibility to ALS (Table 1.1). We also manually added an additional gene, *GLT8D1*, whose variants were recently discovered to cause ALS (Cooper-Knock et al., 2019). All the genes identified so far are involved in remarkably diverse biological roles in the cell including protein quality control and degradation pathways, RNA metabolism, oxidative stress and cytoskeletal dynamics, illustrating the diversity of pathways that are likely to be involved in the disorder (Renton et al., 2014; Taylor et al., 2016). Despite this intricate network, pathogenic variants in only four of these genes contribute to more than 50% of familial cases and more than 10% of sALS, with variants in other genes being found in a small proportion of cases (Renton et al., 2014). These major variants are those affecting *SOD1* (Rosen et al., 1993), the first gene discovered to be associated with ALS, and also *TARDBP* (Kabashi et al., 2008; Sreedharan et al., 2008), *FUS* (Kwiatkowski et al., 2009; Vance et al., 2009), and more recently, *C9orf72* (DeJesus-Hernandez et al., 2011; Renton et al., 2011), which will be the focus of this research project.

Table 1.1. ALS genes. List of genes discovered so far with mutations causally implicated in ALS.

Chromosome	Gene symbol	Gene/locus MIM number	Protein	Inheritance	References
1p36.22	*TARDBP*	605078	TAR DNA-binding protein 43	AD	(Kabashi et al., 2008; Sreedharan et al., 2008)
2p13.1	*DCTN1*	601143	Dynactin subunit 1	AD or AR	(Münch et al., 2004)
2q33.1	*ALS2*	606352	Alsin	AR	(Hadano et al., 2001; Hentati et al., 1994; Yang et al., 2001)
2q34	*ERBB4*	600543	Receptor tyrosine-protein kinase erbB-4	AD	(Takahashi et al., 2013)
2q35	*TUBA4A*	191110	Tubulin alpha-4A chain	AD	(Smith et al., 2014)
3p11.2	*CHMP2B*	609512	Charged multivesicular body protein 2B	AD	(Parkinson et al., 2006)
4q33	*NEK1*	604588	Serine/threonine-protein kinase Nek1	AD	(Kenna et al., 2016)
5q31.2	*MATR3*	164015	Matrin-3	AD	(Johnson et al., 2014)
5q35.3	*SQSTM1*	601530	Sequestosome-1/ Ubiquitin-binding protein p62	AD	(Fecto et al., 2011)
6q21	*FIG4*	609390	Polyphosphoinositide phosphatase	AD	(Chow et al., 2009)

9p21.2	*C9orf72*	614260	Guanine nucleotide exchange C9orf72	AD	(DeJesus-Hernandez et al., 2011; Renton et al., 2011)
9p13.3	*SIGMAR1*	601978	Sigma non-opioid intracellular receptor 1	AR	(Luty et al., 2010)
9p13.3	*VCP*	601023	Transitional endoplasmic reticulum ATPase	AD	(Johnson et al., 2010)
9q34.13	*SETX*	608465	Senataxin	AD	(Chen et al., 2004)
10p13	*OPTN*	602432	Optineurin	AD or AR	(Maruyama et al., 2010)
12q13.12	*PRPH*	170710	Peripherin	AD or AR	(Gros-Louis et al., 2004)
12q13.13	*HNRNPA1*	164017	Heterogeneous nuclear ribonucleoprotein A1	AD	(Kim et al., 2013)
12q13.3	*KIF5A*	602821	Kinesin heavy chain isoform 5A	AD	(Nicolas et al., 2018)
12q14.2	*TBK1*	604834	Serine/threonine-protein kinase TBK1	AD	(Cirulli et al., 2015; Freischmidt et al., 2015)
12q24.12	*ATXN2*	601517	Ataxin-2	AD	(Elden et al., 2010)
14q11.2	*ANG*	105850	Angiogenin	AD	(Greenway et al., 2004, 2006)
15q21.1	*SPG11*	610844	Spatacsin	AR	(Orlacchio et al., 2010)

16p11.2	*FUS*	137070	Fused in Sarcoma	AD or AR	(Kwiatkowski et al., 2009; Vance et al., 2009)
17p13.2	*PFN1*	176610	Profilin-1	AD	(Wu et al., 2012)
18q21	*ALS3*	606640	Unknown	AD	(Hand et al., 2002)
20p13	*ALS7*	608031	Unknown	AD	(Sapp et al., 2003)
20q13.32	*VAPB*	605704	Vesicle-associated membrane protein-associated protein B/C	AD	(Nishimura et al., 2004b, 2004a)
21q22.11	*SOD1*	147450	Superoxide Dismutase 1	AD or AR	(Rosen et al., 1993)
Xp11.21	*UBQLN2*	300264	Ubiquilin-2	XD	(Deng et al., 2011)
22q11.23	*CHCHD10*	615903	Coiled-coil-helix-coiled-coil-helix domain-containing protein 10	AD	(Bannwarth et al., 2014)
22q12.2	*NEFH*	162230	Neurofilament heavy polypeptide	AD or AR	(Figlewicz et al., 1994)
3p21.1	*GLT8D1*	618399	Glycosyltransferase 8 domain-containing protein 1	AD	(Cooper-Knock et al., 2019)

Despite the groundbreaking progress in ALS genetics made in the last three decades and their vital importance for research, these discoveries do not yet have significant clinical relevance, for the following reasons. First, most identified variants have incomplete penetrance. Second,

there is intriguing evidence of several affected families with mutations in more than one of the four major ALS genes. According to Blitterswijk and colleagues (Van Blitterswijk et al.), these cases are statistically more frequent than what would be expected by chance, suggesting a basis of oligogenic inheritance. Finally, despite ongoing gene therapy trials, there are currently no specific therapies for genetic subtypes of ALS. For these reasons, clinical genetic testing is still not very informative and can have serious psychosocial consequences to patients if not done responsibly (Chiò et al., 2014). Nevertheless, experts of the field predict that the increase of genetic data sets available and a greater focus on coding and non-coding rare variants, fueled by lower costs of whole-genome sequencing will continue to drive the ALS genetics field forward and into an increasingly central role, not only in research but also in the clinic (Renton et al., 2014).

1.2 Frontotemporal dementia

1.2.1 Clinical overview of FTD

Frontotemporal dementia (FTD) is a spectrum of neurodegenerative diseases characterised by progressive neurological symptoms encompassing behavioural changes, executive function deficits and language impairments, typically caused by degeneration of the frontal and temporal lobes (Bang et al., 2015). The first detailed report of a patient with FTD is attributed to Arnold Pick, a Czech psychiatrist that described it in 1892 (Pick, 1892), and the disease is now thought to be the second most common form of dementia in people under 65 (Hodges et al., 2003; Sivasathiaseelan et al., 2019).

The FTD spectrum is routinely classified into three main clinical variants according to their leading symptoms in early stages of disease progression. Behavioural-variant FTD (bvFTD) is the most common of these presentations, characterised by progressive changes in behaviour, personality, emotional reactivity and executive skills. The other two variants are designated as primary progressive aphasias (PPAs), with language impairments as the main signs during the

first two years after disease onset. According to the aspects of language that are most affected, PPAs are then classified into non-fluent variant primary progressive aphasia (nfvPPA), characterised by impaired speech and grammar, or semantic variant primary progressive aphasia (svPPA), affecting semantic knowledge and naming (Bang et al., 2015; Olney et al., 2017). Despite the different early presentations, these variants tend to merge at later stages, as the disease affects larger areas of the frontal and temporal lobes. Additionally, the core FTD disorders frequently overlap with other FTD-related traits such as ALS, progressive supranuclear palsy syndrome (PSP-S) and corticobasal syndrome (CBS) (Olney et al., 2017). In the majority of cases, the first symptoms manifest in middle life, between 45 and 64 years of age, although approximately 10% and 30% of patients have an earlier and later disease onset, respectively (Knopman and Roberts, 2011; Snowden et al., 2002). By the end-stage of progression, most FTD patients have overlapping neurological signs from all the different variants and have significant difficulties with basic tasks such as eating, moving and swallowing. Death is usually caused by pneumonia or other infections and tends to happen 6-11 years after disease onset, although certain variants, such as FTD-ALS, are associated with worse survival rates than others (Bang et al., 2015; Hodges et al., 2003).

Diagnosing FTD is a long and complex task that requires a differential diagnosis to rule out a number of overlapping diseases that mirror its symptoms, especially in early stages of progression, and is aided by published standardised guidelines specialising in each main variant (Gorno-Tempini et al., 2011; Rascovsky et al., 2011). The diagnostic process includes careful clinical observation with detailed medical and family history of the patient, as well as neuropsychological tests, neuroimaging, and laboratory testing of blood or even cerebrospinal fluid (CSF). The most common examples of misdiagnosis in FTD include bvFTD patients misdiagnosed with adult-onset psychiatric diseases such as obsessive-compulsive disorder (OCD) and depression, or PPA-linked language impairments being misinterpreted as caused by strokes (Bang et al., 2015; Sivasathiaseelan et al., 2019). Additionally, FTD symptoms are sometimes indistinguishable from Alzheimer's disease (AD), although CSF biomarkers and amyloid imaging can be used to differentiate between the two types of dementia (Olney et al., 2017).

Despite the differences in the initial clinical presentation, FTD variants share the unifying neuropathologic trait commonly designated as Frontotemporal lobar degeneration (FTLD), mainly characterised by neuronal loss and gliosis in the frontal and temporal lobes of the brain (Bang et al., 2015). In addition, similarly to most neurodegenerative diseases,

immunohistochemistry in post-mortem tissue of FTLD cases also reveals intracellular aggregates of different protein composition, allowing the classification of FTLD into five different types of neuropathology (Mackenzie et al., 2009; MacKenzie et al., 2010). In the most common of these, FTLD-TDP, TDP-43 is the main proteinopathy, becoming hyperphosphorylated and ubiquitinated with consequent aggregation in the cytoplasm and, less frequently, in the nucleus (Neumann et al., 2006). The abundance, localization and morphology of the aggregates allow its subclassification into five subtypes (subtypes A, B, C, D and E) (Lee et al., 2017; Mackenzie et al., 2011). Importantly, cases of FTLD-TDP caused by the C9orf72 expansion also display specific histopathology in the form of RNA Foci (DeJesus-Hernandez et al., 2011) or aggregates of different DPR species (Ash et al., 2013; Mori et al., 2013b, 2013a; Zu et al., 2013), in addition to the TDP-43 staining. The second most common FTLD type is FTLD-Tau, in which the microtubule-associated protein tau becomes hyperphosphorylated and deposits abnormally to form aggregates (Hutton et al., 1998; Spillantini et al., 1998). The gene encoding for tau, *MAPT*, can be alternatively spliced, forming tau isoforms with either 3 (3R) or 4 (4R) repeats of the microtubule binding domain. Similarly to FTLD-TDP, FTLD-Tau can also be subdivided into distinct subtypes, according to the tau protein isoform that is predominantly aggregated but also according to their unique immunohistochemical profiles. The most common of these subtypes are Pick's disease, progressive supranuclear palsy (PSP) and corticobasal degeneration (CBD) (Josephs et al., 2011). Approximately 10% of FTLD cases are termed FTLD-FUS, a group of rare diseases sharing the unifying characteristics of FUS as the main aggregated species but also a pronounced caudate atrophy detected post-mortem or by neuroimaging (Josephs et al., 2010; MacKenzie et al., 2011). Similarly to other FTLD types, distinct immunohistochemical patterns allow the subclassification of FTLD-FUS into three disease entities: neuronal intermediate filament inclusion disease (NIFID), basophilic inclusion body disease (BIBD) and atypical FTLD with ubiquitin-only immunoreactive changes (aFTLD-U) (Olney et al., 2017). The fourth and fifth subtypes are the rarest forms of FTLD, and termed FTLD-Ubiquitin-proteasome system (FTLD-UPS) and FTLD no inclusions (FTLD-ni). In FTLD-UPS, inclusions are immunoreactive only to proteins of the UPS such as ubiquitin and P62, and it is still unclear whether another single major component exists (Figure 1-2) (Holm et al., 2007, 2009; MacKenzie et al., 2010). Finally, FTLD-ni encompasses a few cases of FTLD in which no aggregates were observed (Bang et al., 2015; MacKenzie et al., 2010).

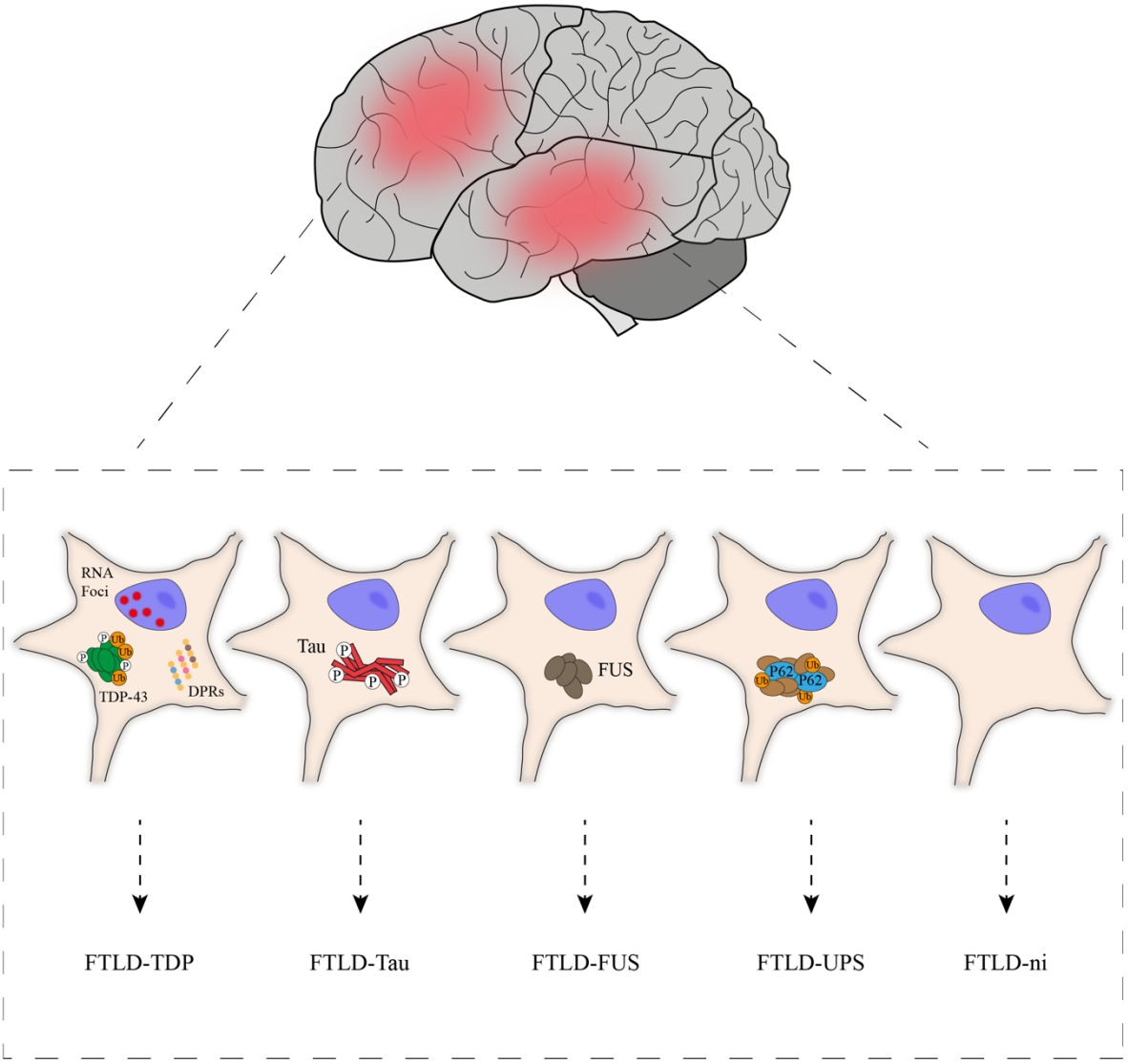


Figure 1-2. The diverse neuropathology of FTD. The varied neuropathological changes associated with FTD are termed Frontotemporal lobar degeneration (FTLD) and typically characterized by marked neurodegeneration in the frontal and temporal lobes of the brain (red areas, upper panel). At the cellular level, FTLD is characterized by the presence of protein aggregates of different species, the composition of which determines its subcategorization (lower panel). In FTLD-TDP, the main aggregated species is TDP-43, hyperphosphorylated and ubiquitinated. In cases of FTLD-TDP caused by the C9orf72 mutation, there is the additional accumulation of RNA Foci and DPRs, more frequently found in the nucleus and cytoplasm, respectively. Hyperphosphorylated Tau and FUS are the main aggregated species in FTLD-Tau and FTLD-FUS, respectively. In cases of FTLD-UPS, the rarest of FTLD subtypes, inclusions are only immunoreactive to components of the UPS, including P62 and Ubiquitin, and it is unclear what other proteins form a part of these aggregates.

To this day, there is no disease-modifying therapy approved for FTD, with patients and their families relying on poorly effective treatments to manage some of the symptoms. Among these, selective serotonin reuptake inhibitors and atypical antipsychotics can improve obsessive-compulsive symptoms and behavioural abnormalities, respectively. Non-pharmacological intervention includes specialized training of caregivers, physical exercise and speech therapy (Bang et al., 2015; Olney et al., 2017).

1.2.2 Epidemiology of FTD

Since the early 2000s, several attempts were made to study the epidemiology of dementias, including FTD diseases, in different regions of the globe (Bernardi et al., 2012; Feldman et al., 2003; Harvey et al., 2003; Ikejima et al., 2009; Knopman and Roberts, 2011; Ratnavalli et al., 2002). Despite these efforts, a worldwide epidemiological perspective on FTD is still an elusive goal, with the most recent numbers still showing significant variability and considered to be underestimations. The reasons for these significant difficulties relate to the significant overlap of FTD with similar disorders, resulting in a high number of misdiagnoses, as well as its low prevalence in the population and the everchanging classification and diagnostic criteria (Bang et al., 2015; Coyle-Gilchrist et al., 2016). A 2014 systematic review of early onset dementia studies (Lambert et al., 2014) estimated the worldwide prevalence of FTD in the population under 65 years of age to vary between 1 and 15.4 per 100,000 people, with the prevalence in Japan being significantly lower than that detected in European studies. However, other studies suggest the numbers in Japan might be much closer to western countries (Wada-Isoe et al., 2012). Researching the same age group, Knopman and Roberts (Knopman and Roberts, 2011) estimated the FTD prevalence in the US to be 15 to 22 per 100,000 people, whereas incidence ranged from 2.7 to 4.1 per 100,000 person-years, showing how the epidemiology of FTD is similar between the European and US population. Finally, expanding research to all age groups, a study in two UK counties (Coyle-Gilchrist et al., 2016) used the revised diagnostic criteria

of the different FTD variants (Gorno-Tempini et al., 2011; Rascovsky et al., 2011) to calculate an FTD prevalence of 10.8 per 100,000 people across all age groups, and an incidence of 1.61 per 100,000 person-years. Importantly, no gender-associated differences were observed in the vast majority of these studies.

Despite these important efforts, robust worldwide epidemiological data of FTD is still lacking, even when compared with equally rare neurodegenerative disorders such as ALS. More studies using the revised classification criteria and without restrictions to young-onset cases are needed in different countries, to understand the global burden of the FTD spectrum (Coyle-Gilchrist et al., 2016).

1.2.3 Genetics of FTD

Similar to many other neurodegenerative diseases, genetic predisposition plays an important role in the aetiology of FTD. In fact, as many as 40% of FTD patients have family history of dementia or psychiatric disorders, with at least 10% having a clear autosomal dominant pattern of inheritance, while the remainder ~60% of cases are generally considered sporadic (Chow et al., 1999; Rohrer et al., 2009). Interestingly, the heritability of FTD is variable across the different clinical variants, with PPAs consistently found to be the least heritable while either bvFTD or FTD-ALS are the most heritable, depending on the cohorts analysed (Goldman et al., 2005; Rohrer et al., 2009).

A search in the OMIM database revealed 12 genes with variants that have been causally associated with FTD (Table 1.2). These are involved in several cellular pathways, the disruption of which is likely to trigger FTD pathogenesis, including proteasomal and autophagy-mediated turnover, the lysosomal/endosomal pathway, DNA/RNA metabolism and cytoskeletal stability, reflecting the complex aetiology of FTD (Pottier et al., 2016). On the reported list (Table 1.2), the majority of genes put together are involved in <5% of cases, but 3 are of particular importance as they are directly implicated in ~60% of all inherited FTD cases (Mishra et al., 2017; Olszewska et al., 2016): *MAPT*, *GRN*, and *C9orf72*, which will be the focus of this project. Despite the remarkable advances of the last 10 years in the molecular

genetics of FTD, there are still a high number of cases with family history of neurodegenerative disorders that are not dominantly inherited, while several of the mutations reported so far have incomplete penetrance. This has prompted FTD researchers to use Genome-wide association studies (GWAS) to identify genetic risk factors that increase the risk of developing the disease and that might modulate the penetrance of established mutations (Van Deerlin et al., 2010; Ferrari et al., 2014; Mishra et al., 2017; Taskesen et al., 2017). Perhaps the best studied of these identified genetic factors is *TMEM106B*, containing single nucleotide polymorphisms (SNPs) that alter its expression levels and consequently modulate the risk of FTLD-TDP-causing mutations in *GRN* and *C9orf72* (Van Deerlin et al., 2010; Nicholson and Rademakers, 2016). It is expected that the future of FTD genetics will focus on exome sequencing and whole-genome sequencing in increasingly large cohorts of sporadic and familial patients to identify new risk factors and genetic modifiers of known mutation carriers (Pottier et al., 2016).

Table 1.2. FTD genes. List of genes discovered so far with variants directly implicated in the aetiology of FTD.

Chromosome	Gene symbol	Gene/locus MIM number	Protein	Inheritance	References
1p36.22	*TARDBP*	605078	TAR DNA-binding protein 43	AD	(Kovacs et al., 2009)
2q35	*TUBA4A*	191110	Tubulin alpha-4A chain	AD	(Smith et al., 2014)
5q35.3	*SQSTM1*	601530	Sequestosome-1/ Ubiquitin-binding protein p62	AD	(Elisa et al., 2012)
9p21.2	*C9orf72*	614260	Guanine nucleotide exchange C9orf72	AD	(DeJesus-Hernandez et al., 2011; Renton et al., 2011)

9p13.3	*VCP*	601023	Transitional endoplasmic reticulum ATPase	AD	(Watts et al., 2004)
12q14.2	*TBK1*	604834	Serine/threonine-protein kinase TBK1	AD	(Freischmidt et al., 2015)
14q24.2	*PSEN1*	104311	Presenilin-1	AD	(Raux et al., 2000)
16p11.2	*FUS*	137070	Fused in Sarcoma	AD	(Van Langenhove et al., 2010; Yan et al., 2010)
17q21.31	*GRN*	138945	Progranulin	AD	(Baker et al., 2006; Cruts et al., 2006)
17q21.31	*MAPT*	157140	Microtubule-associated protein tau	AD; Multifactorial	(Hutton et al., 1998)
22q11.23	*CHCHD10*	615903	Coiled-coil-helix-coiled-coil-helix domain-containing protein 10	AD	(Bannwarth et al., 2014)
Xp11.21	*UBQLN2*	300264	Ubiquilin-2	XD	(Deng et al., 2011)

1.3 C9orf72-linked ALS/FTD

1.3.1 The hexanucleotide repeat expansion of C9orf72

The discovery of the C9orf72 mutation as a common genetic cause for both ALS and FTD was a decades-long endeavor, involving several research groups and seminal findings. As mentioned above, while some cases of ALS do spare cognitive function, there were early studies reporting that up to 50% of patients ended up suffering from symptoms that include personality changes and impaired executive function consistent with an FTD spectrum disease (Ferrari et al., 2011; Lomen-Hoerth et al., 2003; Strong et al., 1999). These observations were strengthened by neuroimaging studies identifying frontal and temporal lobe anomalies in ALS patients, suggesting a clinical overlap between both disorders (Abrahams et al., 1996; Strong, 2008; Talbot et al., 1995). In 2006, a pivotal study by Neumann and colleagues (Neumann et al., 2006) uncovered a common neuropathological hallmark to both diseases, in which TDP-43 is the main component of ubiquitinated inclusions in CNS tissue from both FTD and ALS patients, while being apparently absent in aggregates of other neurodegenerative traits (Arai et al., 2006; Neumann et al., 2006). In addition to the clinical and neuropathological overlap, reports of families with relatives having either ALS, FTD or both, in what appeared to be a Mendelian dominant pattern of inheritance, suggested a common genetic origin for both disorders as well (Gunnarsson et al., 1991; Hosler et al., 2000; Majoor-Krakauer et al., 1994). A series of genetic linkage studies and the advent of Next-generation sequencing (NGS) traced this genetic origin to several loci (Gijselinck et al., 2010; Hosler et al., 2000), with the most common one locating in chromosome 9p (Morita et al., 2006; Vance, 2006), although the exact gene and the nature of the mutation involved still eluded experts. Finally, in 2011, two independent groups identified a hexanucleotide repeat expansion (GGGGCC or G4C2) in the intronic region of the C9orf72 gene as the most common genetic cause for familial and sporadic cases of both ALS and FTD (DeJesus-Hernandez et al., 2011; Renton et al., 2011). Indeed, a cross-sectional study from 17 regions worldwide estimated that the C9orf72 mutation is present in ~37% and ~6% of familial and sporadic ALS cases, respectively, as well as ~25% and ~6% of familial and sporadic cases of FTD, respectively (Majounie et al., 2012). Importantly, though, this study also detected significant geographical variation of this mutation, possibly

due to founder effects, being common in patients from Europe or of European ancestry but very rare in Asian populations (Majounie et al., 2012).

The exact number of G4C2 repeats above which an individual develops ALS and/or FTD remains unclear, as individuals with small expansion sizes in the range of 20-30 repeats can either develop the disease or remain asymptomatic throughout a normal life span (Balendra and Isaacs, 2018). Nevertheless, the vast majority of the population has up to 10 repeats of the G4C2 sequence, while most patients often possess up to hundreds or even thousands (Beck et al., 2013). Importantly, the number of repeats detected in a certain individual should be interpreted with caution, as the intrinsic instability of the G4C2 repeat expansion frequently results in somatic mosaicism, observed between different cell types but also within the same tissue (van Blitterswijk et al., 2013; Ebbert et al., 2018; Nordin et al., 2015; Suh et al., 2015). This characteristic might help to explain the age-dependent incomplete penetrance of this mutation, with some expansion carriers being asymptomatic even in their ninth decade of life, which raises significant challenges to the design of clinical trials and genetic counselling (Majounie et al., 2012; Murphy et al., 2017).

According to the ENSEMBL database (Yates et al., 2020), the C9orf72 gene is ~38 kilo bases (Kb)-long and spans a total of 11 exons (10 of which are protein coding), with the G4C2 repeat region located between the non-coding exons 1a and 1b (Figure 1-3). The gene is alternatively spliced, forming 3 main annotated transcripts (V1, V2 and V3) which encode for either a long (C9orf72-L) or a short (C9orf72-S) protein isoform (Balendra and Isaacs, 2018; Haeusler et al., 2016). Both the transcripts V2 and V3 encode for the long isoform, with 481 amino acids (aa), although they are regulated by different promoters. Indeed, while V3 is regulated by a promoter located within exon 1a, V2 expression is driven by an alternative endogenous promoter located in the first intron of the C9orf72 gene (DeJesus-Hernandez et al., 2011). With 222 aa, C9orf72-S is encoded by transcript V1, which is regulated by the same promoter as V3, and contains the first 5 exons of the gene, 4 of which are protein coding (Figure 1-3).

1.3.2 The biological role of C9orf72

The ubiquitous cellular function of the C9orf72 protein was largely unknown for a long time, even after the discovery of the pathogenic expansion, until several studies divulged compelling evidence for a central role in the regulation of vesicular trafficking. This role has been studied in particular detail in the context of macroautophagy (herein termed autophagy) (Webster et al., 2016a). Being crucial for cell homeostasis and survival, autophagy is a recycling process in which damaged organelles and misfolded proteins are transported to the lysosome for degradation, and its dysfunction has been associated with numerous neurodegenerative diseases, including other subtypes of ALS (Mizushima and Komatsu, 2011; Ramesh and Pandey, 2017). The first clues for this role of C9orf72 protein came from bioinformatics analysis, identifying a homolog of the 'differentially expressed in normal and neoplastic cells' (DENN) module, a group of conserved domains that confer the characteristics of GDP-GTP exchange factors (GEFs) (Levine et al., 2013; Zhang et al., 2012). Proteins with DENN domains are able to activate Rab GTPases by inducing their dissociation of Guanosine diphosphate (GDP) and consequent binding to Guanosine triphosphate (GTP). In this way, DENN-containing proteins can regulate several Rab-dependent pathways mainly related to endocytosis and membrane trafficking, including autophagy (Marat et al., 2011). These bioinformatic predictions were corroborated experimentally, showing that C9orf72 co-localized with several members of the Rab family and co-migrated with lysosomal vesicles, while its depletion dysregulated autophagy in cell lines and primary neurons (Farg et al., 2014). Expanding on these results, several independent teams then gradually uncovered a complicated web of interactions and biological functions of C9orf72, exposing its increasingly central role in regulating different aspects of the autophagy-lysosomal pathway (Figure 1-3). Indeed, Webster and colleagues (Webster et al., 2016b) showed that C9orf72-L and, to a lesser extent, C9orf72-S, were crucial in the translocation of the Unc-51-like kinase 1 (ULK1) initiation complex to the phagophore by acting as an effector for Rab1a. In parallel, other groups first reported how C9orf72-L formed a complex with Smith-Magenis Syndrome Chromosome Region candidate 8 (SMCR8) and WD repeat-containing protein 41 (WDR41) (Sellier et al., 2016; Sullivan et al., 2016; Xiao et al., 2016; Yang et al., 2016). The C9orf72-SMCR8-WDR41 complex was shown to also regulate the initiation of autophagy through its interaction with the ULK1 complex and was shown to act as a GEF for Rab8a and Rab39b, regulating in this way

the autophagic flux (Sellier et al., 2016; Sullivan et al., 2016; Yang et al., 2016). The GEF activity of C9orf72-SMCR8-WDR41, however, is still not universally accepted, as recent work now suggests that the complex might actually function as a GTPase activating protein (GAP) for Rabs, namely Rab8a and Rab11a (Tang et al., 2020b, 2020a). Despite the uncertainty still surrounding the precise function of this complex, the key role of C9orf72 in autophagy is now undeniable and was also demonstrated *in vivo*. Indeed, C9orf72 knockout (KO) mice do not develop ALS or FTD, but show autoimmune and inflammation phenotypes similar to what has been observed in autophagy-deficient mice (Atanasio et al., 2016; Burberry et al., 2016; Koppers et al., 2015; O'Rourke et al., 2016; Sudria-Lopez et al., 2016; Sullivan et al., 2016). Also, autophagy and lysosomal proteins such as Microtubule-associated protein 1 light chain 3 (LC3), Lysosomal associated membrane protein 1 (LAMP1) and P62 were found to accumulate in the liver and spleen of some of these animals. This reinforced the idea that the aberrant macrophage and microglial function observed might result from dysfunctional autophagy (O'Rourke et al., 2016; Sullivan et al., 2016). The biological function of C9orf72 is not strictly restricted to autophagy, though, having been shown to also regulate endosomal trafficking, while lowering its expression levels reduces endocytosis (Farg et al., 2014). Expanding on this, Aoki and colleagues (Aoki et al., 2017) have demonstrated that this regulation is at least partially dependent on the interaction with Rab7L1, and the KO of C9orf72 in mice reinforced this role as it resulted in impaired endosomal trafficking *in vivo* (Figure 1-3) (O'Rourke et al., 2016).

Therefore, significant evidence now exists that C9orf72-L is required for normal intracellular vesicle transport and is particularly important in the regulation of endocytosis and the autophagy-lysosomal pathway. The shorter C9orf72-S isoform, despite having also been implicated in this pathway (Webster et al., 2016b), has been considerably less studied than the long isoform, and was shown to locate in the nuclear membrane, possibly regulating nucleocytoplasmic transport (Figure 1-3) (Xiao et al., 2015).

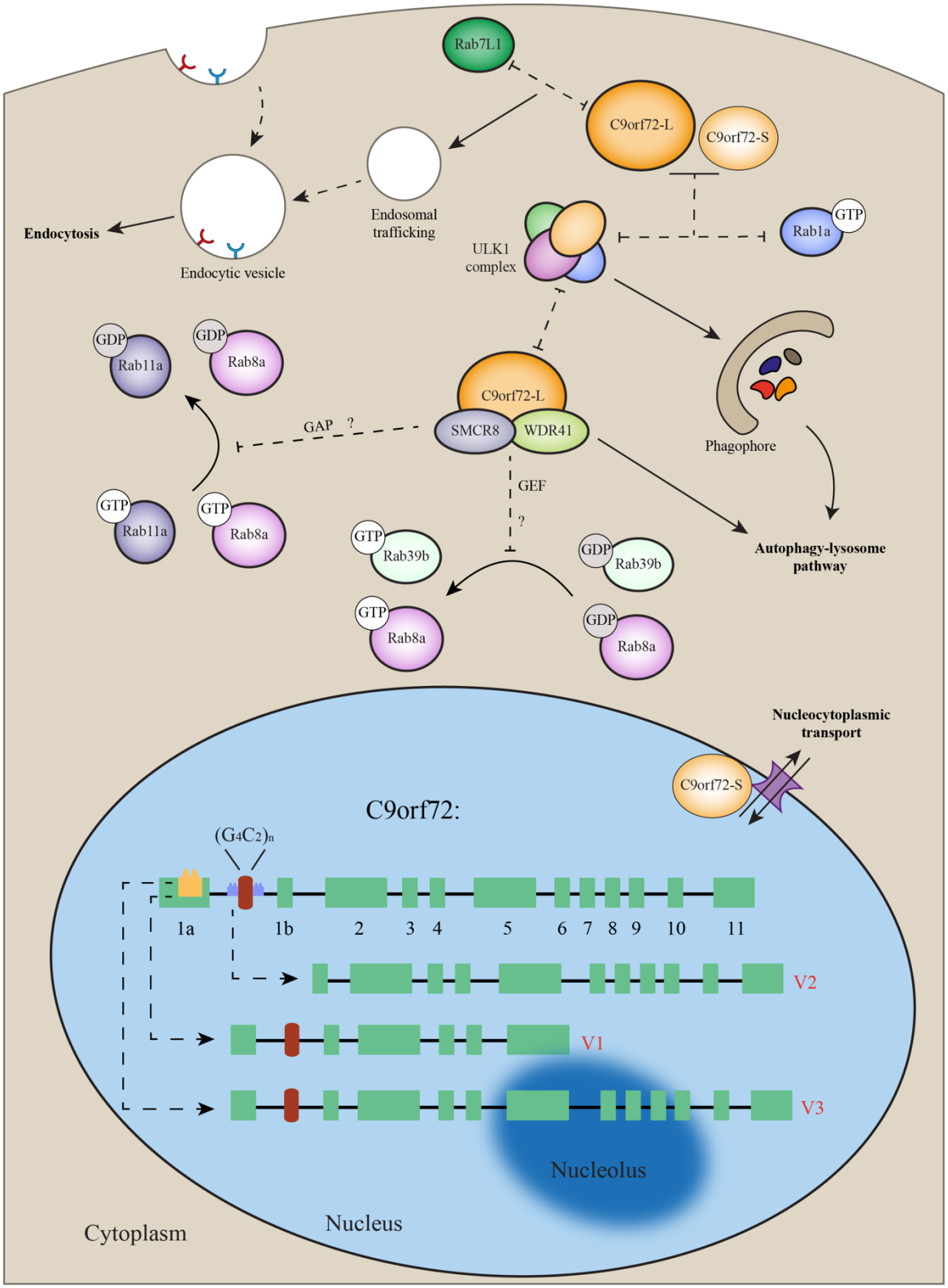

Rab7L1
C9orf72-L
C9orf72-S
Endocytosis
Endocytic vesicle
Endosomal trafficking
ULK1 complex
GTP
Rab1a
GDP
Rab11a
GDP
Rab8a
C9orf72-L
SMCR8
WDR41
GAP ?
Phagophore
GEF
?
GTP
Rab11a
GTP
Rab8a
GTP
Rab39b
GDP
Rab39b
GTP
Rab8a
GDP
Rab8a
Autophagy-lysosome pathway
Nucleocytoplasmic transport
C9orf72-S
C9orf72:
(G4C2)n
1a
1b
2
3
4
5
6
7
8
9
10
11
V2
V1
V3
Nucleolus
Cytoplasm
Nucleus

Figure 1-3. The genetics and cellular functions of C9orf72. Schematic representation of the C9orf72 gene and its transcripts, and the main biological roles that the protein product plays in the cell. The C9orf72 gene is represented in the nucleus, with a length of approximately 38 Kb spanning a total of 11 exons, 10 of which are protein coding (not to scale), and is regulated by two different promoters. Depicted in yellow, the more upstream promoter is located within exon 1a and drives expression of both transcripts V1 and V3 through alternative splicing, whose sequence contain the G4C2 repeat region. The smaller V1 transcript encodes for the shorter isoform, C9orf72-S, with 222 aa, while the V3 encodes for the larger isoform, C9orf72-L, with 481 aa. Alternatively, transcription of the gene can be driven by an endogenous promoter located within the first intron of the C9orf72, depicted in blue, driving the expression of transcript V2, which also encodes for C9orf72-L.

In the cytoplasm, a schematic representation is shown that depicts the main cellular pathways that the C9orf72 protein is involved in, namely the autophagy-lysosome pathway, endocytosis and possibly nucleocytoplasmic transport. Both C9orf72 isoforms regulate the initiation of autophagy through their interaction with Rab1a and the ULK1 initiation factor, while C9orf72-L forms a complex with SMCR8 and WDR41. This complex regulates the autophagic flux by acting as a GEF or a GAP to regulating Rab proteins. C9orf72-L was also shown to regulate endocytosis, at least partly by interacting with Rab7L1 and regulating endosomal trafficking. C9orf72-S is a less studied isoform that was shown to locate to the nuclear membrane, where it is thought to regulate nucleocytoplasmic transport.

1.3.3 Pathophysiological mechanisms of C9orf72

The molecular mechanisms through which the G4C2 expansion in the C9orf72 triggers neurodegeneration has been the subject of intense studies and debate since its discovery (Balendra and Isaacs, 2018; van Blitterswijk et al., 2012; Gendron and Petrucelli, 2018; Gendron et al., 2014; Gitler and Tsuiji, 2016; Haeusler et al., 2016; Heutink et al., 2014; Mizielinska and Isaacs, 2014; Rohrer et al., 2015). Three main mechanisms have been proposed that have garnered substantial support, including a loss of protein function and the gain-of-function mechanisms associated with expanded RNA toxicity and the accumulation of DPRs.

Suspicions about a loss of function mechanism arose from early qPCR experiments reporting a significant reduction at the transcript level in C9orf72 patients (DeJesus-Hernandez et al., 2011; Gijselinck et al., 2012), which was later validated at the protein level through western-blot of post-mortem frontal cortex tissue (Waite et al., 2014). The loss-of-function hypothesis was also supported by experiments on simple animal models, including a gene KO and knockdown in *C. elegans* (Therrien et al., 2013) and zebrafish (Ciura et al., 2013) respectively, which induced behavioral deficits and motor neuron pathology. In addition to this, reduced protein levels of C9orf72 in human induced motor neurons was shown to be sufficient to trigger neurodegeneration levels comparable to that of patient-derived cells (Shi et al., 2018). Considering these results in light of the important regulatory role of C9orf72 in vesicle trafficking and autophagy, as well as the frequent mutations in other autophagy-associated genes causing ALS (Mizushima and Komatsu, 2011; Ramesh and Pandey, 2017), it was plausible that loss-of-function could be the predominant mechanism (Figure 1-4). However, conflicting evidence has been presented that demonstrates how loss of function alone is not sufficient to explain many key features of the C9orf72-linked ALS/FTD. Among this evidence is the recognition that homozygous human patients do not display more severe symptoms than heterozygous (Fratta et al., 2013) as well as data from several mouse models of C9orf72 loss-of-function. Indeed, in these studies, reduction or complete ablation of mouse C9orf72 levels was insufficient to trigger neurodegeneration characteristic of ALS/FTD, despite some of these models presenting with inflammation and autoimmune phenotypes linked to autophagy defects (Atanasio et al., 2016; Burberry et al., 2016; Jiang et al., 2016; Koppers et al., 2015; Lagier-Tourenne et al., 2013; O'Rourke et al., 2016; Sudria-Lopez et al., 2016; Sullivan et al., 2016; Ugolino et al., 2016). The disparity of results obtained in mice when comparing with simpler animal models can be explained by the considerably higher degree of homology between human and murine C9orf72 than that of orthologues from zebrafish and *C. elegans* (Haeusler et al., 2016). Taking all this into consideration, loss-of-function is not likely to be the primary pathogenic mechanism of C9orf72-linked neurodegeneration, and gain-of-function paradigms must also be taken into account.

RNA-mediated toxicity was proposed as a possible pathogenic mechanism of C9orf72-linked neurodegeneration in the very first reports of the hexanucleotide expansion (DeJesus-Hernandez et al., 2011; Renton et al., 2011), with authors showing evidence of repeat-expanded RNA forming nuclear Foci in post-mortem patient tissue. These RNA Foci were also shown to accumulate in disease models, including induced pluripotent stem cells (iPSC)-derived neurons

from affected individuals (Almeida et al., 2013; Dafinca et al., 2016; Donnelly et al., 2013; Sareen et al., 2013). Considering the well-studied toxicity of RNA derived from other non-coding pathogenic repeat expansions (Li and Jin, 2012; Udd and Krahe, 2012), it was hypothesized that the accumulation of expanded C9orf72 transcripts can also hijack multiple RNA-binding proteins (RBPs), including key splicing components. This can then lead to aberrant mRNA splicing and processing, ultimately resulting in a global dysregulation of cellular pathways (Figure 1-4) (Todd and Paulson, 2010). Expanding on this hypothesis, a series of different studies identified multiple RNA-binding proteins which were shown to interact with C9orf72 transcripts both *in vitro* and in patient's tissue, with some of these hits confirmed to co-localize with RNA Foci by immunohistochemistry (Cooper-Knock et al., 2014a; Donnelly et al., 2013; Haeusler et al., 2014; Mori et al., 2013c; Sareen et al., 2013; Xu et al., 2013). Although they differed according to the study, overlapping all the interactomes identified the protein family of heterogeneous nuclear ribonucleoproteins (hnRNPs), involved in splicing and gene expression (Geuens et al., 2016), as an overrepresented interactor, strengthening the RNA-RBP hypothesis (Haeusler et al., 2016). Also, in addition to aberrant splicing and nuclear stress, impaired nucleocytoplasmic transport is another toxic consequence of expanded C9orf72 transcripts (Figure 1-4). Indeed, RanGAP1, crucial for the active nucleocytoplasmic shuttling of nuclear localization signal (NLS)-containing proteins, was also identified as a strong interactor of G4C2 RNA and as a potent suppressor of C9orf72-mediated toxicity (Donnelly et al., 2013; Zhang et al., 2015). Despite this mounting evidence, comparative studies in *Drosophila melanogaster* (Mizielinska et al., 2014; Tran et al., 2015) as well as mouse models created with a bacterial artificial chromosome (BAC) (O'Rourke et al., 2015; Peters et al., 2015) do not support a key toxicity role for G4C2-containing repeat RNA. Indeed, these animal models successively reproduce the appearance of RNA Foci while consistently failing to induce motor phenotype or neurodegeneration. Although a recent BAC mouse model with ~500 hexanucleotide repeats reports severe motor impairments, neuronal loss and decreased survival alongside Foci accumulation (Liu et al., 2016), the previous studies highlight that the repeat RNA alone is not sufficient to trigger ALS/FTD.

Repeat associated non-ATG-translation (RAN) was first described as a new pathophysiological pathway in Spinocerebellar ataxia type 8 (SCA8) and consists of a non-canonical mechanism in which a nucleotide repeat expansion can be translated in the absence of an ATG start codon. In SCA8, this meant that the CAG expansion of the *ATXN8* gene is translated into toxic polyserine and polyalanine peptides in addition to the canonical polyglutamine protein (Zu et

al., 2011). In C9orf72-ALS/FTD, this mechanism was first identified by two independent groups that reported the aggregation of RAN dipeptide repeat proteins (DPRs) in patient's tissue, consisting of a long peptide stretch containing two alternating amino acids: poly-(GP); poly-(GA); and poly-(GR) (Ash et al., 2013; Mori et al., 2013a). These aggregated repeat peptides are, with some exceptions, mostly cytoplasmic and are the result of translation in all the three reading frames of the C9orf72 repeat expansion, without an ATG start codon, which is consistent with the definition of RAN translation. Later on, several groups also described the presence of DPRs derived from antisense RNA molecules (a product of bidirectional transcription) in post-mortem tissue from patients: poly-(PA); poly-(PR); and poly-(GP), which is translated from both the sense and antisense strands (Gendron et al., 2013; Mann et al., 2013; Mori et al., 2013b; Zu et al., 2013). These studies generated tremendous excitement for the possibility that DPRs might be the main driver of neurodegeneration in C9orf72-ALS/FTD, a hypothesis backed by a multitude of experiments in animal models. Indeed, many different research groups worldwide used models of zebrafish (Ohki et al., 2017; Swaminathan et al., 2018), *D. melanogaster* (Freibaum et al., 2015; Mizielinska et al., 2014; Xu and Xu, 2018) and mice (Choi et al., 2019; Schludi et al., 2017; Zhang et al., 2018) to show how expression of DPRs, particularly arginine-containing species, can cause neurodegeneration, in many cases resulting in various behavioural deficits. These and other studies also uncovered a remarkably complex network of pathogenic pathways elicited by the different DPRs that are behind this toxicity. The most commonly found and aggregation-prone, poly-(GA), is known to sequester several important protein targets, of which the most common are components of the UPS such as HR23A and HR23B (Zhang et al., 2016). Possibly through this sequestering, poly-(GA) DPRs cause proteasomal inhibition and, consequently, accumulation of misfolded proteins in the Endoplasmic reticulum (ER), also causing ER stress (Figure 1-4). Interestingly, ER stress inhibitors protect against poly-(GA)-induced neurotoxicity (Zhang et al., 2014). Expanding on the consequences of poly(GA)-mediated sequestration, the transport factor Unc119, involved in neuromuscular and axonal function, was also found to bind poly-GA DPRs. This sequestering was shown to cause loss-of-function of Unc119, which in turn inhibits dendritic branching and causes neurotoxicity, while restoring Unc119 expression partially rescues poly-(GA)-induced toxicity (Figure 1-4) (May et al., 2014). Interestingly, poly-(GA) can also sequester proteins involved in nucleocytoplasmic transport, a major mechanism whose disruption has been frequently associated with C9orf72-mediated neurodegeneration (Freibaum et al., 2015; Zhang et al., 2015, 2016). In addition to these proposed pathways, poly-(GA) DPRs were also shown to induce R-loop formation and cause defects in the ATM-

mediated DNA repair signalling, precipitating C9orf72-mediated DNA damage *in vitro* and *in vivo* (Figure 1-4) (Walker et al., 2017). Despite the important pathogenic mechanisms uncovered for poly-(GA), the most toxic species of DPRs are still considered to be arginine-containing ones, namely poly-(GR) and poly-(PR). In two elegant studies, these were shown to bind to RBPs and proteins with low-complexity domains (LCDs), which regulate membrane-less organelles such as nucleoli, nuclear pore complexes and stress granules (Lee et al., 2016; Lin et al., 2016). Importantly, these DPRs were shown to alter the liquid-liquid phase separation (LLPS) properties of the LCD-containing proteins, disrupting in this way the function and dynamics of these membrane-less organelles (Lee et al., 2016; Lin et al., 2016). To the affected cells, the toxic consequences of this disruption are numerous, including impaired ribosome biogenesis, altered nucleocytoplasmic transport, and a global dysregulation of translation (Figure 1-4) (Freibaum and Taylor, 2017). Interestingly, poly-(GR) and poly-(PR) DPRs were also shown to trigger DNA damage in mammalian cells through R-loop-independent mechanisms (Figure 1-4) (Farg et al., 2017; Lopez-Gonzalez et al., 2016). As for the remaining DPRs, poly-(GP) and poly-(PA), these were found to be non-toxic, although the biochemical characteristics of poly-(GP) make it a powerful disease biomarker with important implications for monitoring therapeutic benefit in future clinical trials (Balendra et al., 2017; Gendron et al., 2017; Lehmer et al., 2017). Despite the substantial data pointing towards a major role of DPRs in the pathophysiology of C9orf72, there are arguments against this interpretation. One of the issues that warrants consideration is that the majority of these studies use overexpression paradigms, which is likely to result in expression levels higher than what would be found in patients. Expanding on this, models in which expression levels are actually more physiological, such as some of the BAC transgenic mice, display DPRs without an obvious motor phenotype or neurodegeneration (O'Rourke et al., 2015; Peters et al., 2015). In addition to this, post-mortem observations that most DPR inclusions do not co-localise with neurodegeneration, excellently reviewed elsewhere (Balendra and Isaacs, 2018), also cast doubt over the actual severity of DPR toxicity in C9orf72-ALS/FTD patients.

Finally, studies in iPSC-derived neurons from patients also revealed dysregulated Ca^{2+} signalling as a pathogenic mechanism linked to the C9orf72 expansion as well as *TARDBP* mutations, caused by altered stoichiometry of proteins regulating the mitochondrial Ca^{2+} uniporter (MCU) (Dafinca et al., 2016, 2020). This results in decreased expression of the MCU and impaired mitochondrial Ca^{2+} buffering from the cytosol as a consequence (Figure 1-4). Nevertheless, it is still unclear whether this mechanism results from loss or gain-of-function

consequences and whether it might be driven by expression of RNA Foci or DPRs (Dafinca et al., 2016, 2020).

All in all, recent years have witnessed tremendous progress in uncovering the pathogenicity of C9orf72-linked ALS/FTD and the evidence gathered so far suggests that single, isolated mechanisms are insufficient to trigger disease. Instead, the loss and gain-of-function events synergising with ageing are likely to be co-conspirators in driving the neurodegeneration characteristic of these diseases. Recently, *in vitro* and *in vivo* evidence has been put forward that convincingly supports this hypothesis, demonstrating how the consequences of C9orf72 haploinsufficiency exacerbate gain-of-function mechanisms (Boivin et al., 2020; Shi et al., 2018; Zhu et al., 2020).

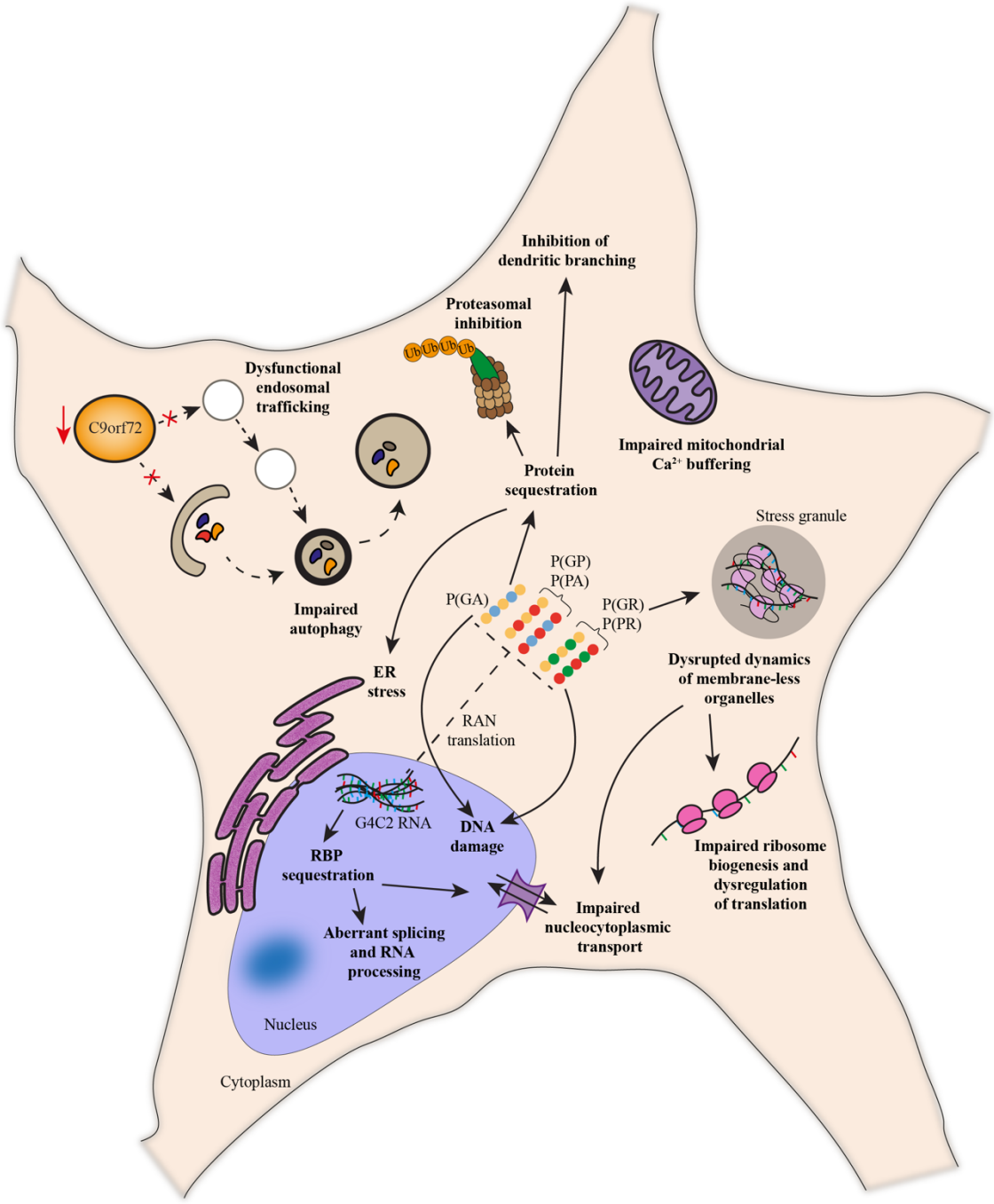


Figure 1-4. The pathophysiology of C9orf72-linked neurodegeneration. Schematic representation illustrating the pathological mechanisms associated with C9orf72 neurodegeneration. Haploinsufficiency leading to reduced C9orf72 expression can result in dysfunctional endosomal trafficking and impaired autophagy, exacerbating gain-of-function mechanisms. Accumulation of G4C2 expansion-containing transcripts can sequester RBPs, resulting in aberrant splicing and RNA metabolism, as well as impaired nucleocytoplasmic transport. RAN translation of these transcripts can

form 5 different species of DPRs, of which poly-(GP) and poly-(PA), labelled as P(GP) and P(PA), are not considered to be toxic. The highly aggregation-prone Poly-(GA), labelled as P(GA), can induce DNA damage in addition to sequestering protein targets, which has toxic consequences ranging from proteasomal inhibition, ER stress and inhibition of dendritic branching. Poly-(GR) and poly-(PR), labelled as P(GR) and P(PR), can cause DNA damage and disrupt the dynamics of membrane-less organelles such as stress granules, but also nucleoli and nuclear pore complexes. This results in toxicity by impairing crucial cellular mechanisms such as ribosome biogenesis, regulation of translation and nucleocytoplasmic transport. In addition, or in parallel, to these pathways, dysregulated Ca^{2+} signalling is another major pathogenic mechanism identified in C9orf72- and TARDBP-associated neurodegeneration, at least partly caused by impaired Ca^{2+} buffering by the mitochondria.

1.3.4 Small molecule- and antibody-based therapies for C9orf72-ALS/FTD

As mentioned above, ALS and FTD patients suffer from devastating diseases but the list of therapeutic options offered to them is still dramatically inefficient in slowing disease progression or extending their life expectancy, despite decades of research. This situation reflects the complex aetiology of these disorders and highlights the difficulty in developing a global treatment that will benefit patients across the entire spectrum of ALS and FTD. Uncovering the pathophysiology of these diseases, however, allows the development of more targeted therapies that might prove more efficacious. Although only 10 years have passed since the discovery of the G4C2 expansion of the C9orf72 gene (DeJesus-Hernandez et al., 2011; Renton et al., 2011), numerous treatments have been developed that target either the expansion itself, its transcriptional and translational by-products or its downstream mechanisms (Balendra and Isaacs, 2018; Mis et al., 2017). Small molecules that disrupt the G-quadruplex structures formed by expanded RNA have been successful in preventing sequestration of RBPs, decreasing RNA Foci and DPR accumulation as well as alleviating toxicity (Simone et al., 2018; Su et al., 2014; Zamiri et al., 2014; Zhang et al., 2015). Alternatively, small molecules that inhibit the ER stress frequently associated with C9orf72 neurodegeneration have also shown promise in ameliorating poly-(GA) toxicity in primary neurons (Zhang et al., 2014). Recently, the Food and Drug Administration (FDA)-approved drug metformin was reported to decrease DPRs and improve behavioural and neuropathological aspects of C9orf72-ALS/FTD

in vivo by inhibiting RNA-dependent protein kinase (PKR) (Zu et al., 2020). In addition to these, passive and active immunization are other approaches that have shown benefit in other neurodegenerative diseases characterised by protein aggregation such as AD, Parkinson's disease (PD) and FTLD-Tau (Valera and Masliah, 2013). In C9orf72-ALS/FTD, poly-(GA)-specific antibodies were successful in reducing the accumulation and spreading of these DPRs *in vitro*, while its use in transgenic mice reduced microglial activation, neurodegeneration and improved behavioural aspects of the models (Nguyen et al., 2020; Zhou et al., 2017, 2020). Despite encouraging results with small molecules and immunotherapy, the majority of the treatments developed for C9orf72-ALS/FTD so far, including the ones closest to clinical application, have been gene therapies and will be discussed in the next section.

1.4 Gene therapy

1.4.1 Gene therapy vectors for neurodegenerative diseases

Gene therapy is the term conferred to a broad group of treatments in which nucleic acids are used as the therapeutic molecule. It generally consists in expressing a transgene to restore a defective protein or express therapeutic genes that intervene in the disease mechanism. Alternatively, gene therapy can also involve the silencing or editing of a toxic allele using molecular techniques such as RNA interference and gene editing (Anguela and High, 2019). Since it was first idealized, gene therapy has captivated the imagination of scientists due to its potential to treat or even cure genetic disorders, for which symptom management and palliative care are often the only course available to patients (Goswami et al., 2019). Despite considerable setbacks throughout the years, the development and optimization of more reliable and efficient delivery systems in recent years has resulted in some successful clinical trials and brought the topic back to the spotlight of the scientific community (Naldini, 2015). Indeed, the clinical benefit of gene therapy holds tremendous promise, with potential treatments being constantly developed for a wide array of diseases including, but not limited to, cardiac diseases, cancer, hepatic disease and neurodegenerative disorders, which is the focus of this project (Ishikawa

et al., 2018; O'Connor and Boulis, 2015; Piccolo and Brunetti-Pierri, 2015; Saadatpour et al., 2016). Indeed, neurodegenerative diseases, particularly those with a defined genetic mutation at its origin, are ideal targets for gene therapy development since they usually have few therapeutic alternatives and there is striking potential for alleviating the suffering of patients. A remarkable example of this potential is the significant therapeutic benefit of Zolgensma, a gene therapy treatment which was successful in slowing down the progression of spinal muscular atrophy (SMA) (Mendell et al., 2017). In this trial, treated children had dramatically improved survivability upon a single intravenous injection and were able to reach neurological and developmental milestones that would have been otherwise impossible, leading to the recent approval of this therapy for clinical use (Hoy, 2019; Mendell et al., 2017).

A crucial step when developing a new gene therapy is choosing the most suitable delivery system to allow adequate expression of the transgene in the target population of cells while keeping immunogenicity and off-targets to a minimum (Maguire et al., 2014). The available vectors for the nervous system are commonly grouped in one of two categories, non-viral and viral, each one having important advantages and downsides.

As the name implies, non-viral gene therapy consists in administering nucleic acids as a therapeutic strategy without relying on viruses. This can mean either the administration of naked DNA/RNA such as antisense oligonucleotides (ASOs), or their combination with various biomaterials to aid in the transport across the cellular membrane, among which the most common are lipid-based nanoparticles, polymeric materials, surfactants and carbon nanostructures (Foldvari et al., 2016). Among these candidates, ASOs have proven to be particularly promising for neurodegenerative disorders, with an FDA-approved product to treat SMA and several other treatments having reached human clinical trials for ALS, SMA and Huntington's disease (HD) (Schoch and Miller, 2017; Talbot and Wood, 2019). Delivered into the CSF through intrathecal or intraventricular administration, ASOs can modulate gene expression in the CNS, either by inducing mRNA degradation via RNAse H recruitment, modifying alternative splicing or by microRNA (miRNA) sequestration (Schoch and Miller, 2017). In addition to this, naked molecules and non-viral vectors in general have a high safety profile and affordable costs of production. These successes and advantages, however, are counterbalanced by a decreased transfection efficiency when compared to viral vectors (Ingusci et al., 2019; Yin et al., 2014a), as well as the need for sustained administration. The challenges associated with the low transfection are varied and include the degradation of the

nucleic acids by endonucleases in the extracellular space, but also difficulties related to endosomal escape, vector unpacking and transport to the nucleus (Yin et al., 2014a).

Viral gene therapy harnesses the natural capability of viruses to introduce and express their genomic information in the nucleus of infected cells. Several types of viral vectors are currently used to develop gene therapies for neurodegenerative diseases including adeno-associated viruses (AAVs), retroviruses, lentiviruses, adenoviruses and Herpes simplex virus (HSV) (Jayant et al., 2016). Of these, lentiviruses and AAVs have been the undeniable frontrunners due to their efficiency and low toxicity, which led to their wide usage in both pre-clinical and clinical studies (Choudhury et al., 2016). Indeed, not only can they transduce dividing and non-dividing cells with high efficiency, but they also allow for a robust expression of the transgene that can last for years or even decades (Choudhury et al., 2016). Lentiviruses were the chosen vectors in numerous pre-clinical and some clinical studies to treat ALS, SMA, HD and PD thanks to their low immunogenicity, reasonable transgene packaging capacity (~8 kb) and simplicity in producing high titers (Nanou and Azzouz, 2009). As disadvantages, lentiviruses have a low volumetric spread upon injection, limited to 400-600 µm, which results in fewer cells being transduced, and they integrate their genetic material in the host genome, with oncogene activation as an uncommon but daunting consequence (Kotterman et al., 2015; Osten et al., 2006). In contrast, recombinant AAVs have a wider volumetric spread (1-3 mm) (Kells et al., 2009) and the transgene is kept in an episomal state rather than being integrated in the host genome, bypassing the potential complications of insertional mutagenesis associated with lentiviruses (Mulcahy et al., 2015; Murlidharan et al., 2014). Also, despite concerns of adverse inflammatory reactions resulting from AAV delivery in non-human primates (NHP) and humans (Hinderer et al., 2018; Mueller et al., 2020), these are generally well managed with adequate immunosuppression, and AAVs continue to maintain a good safety profile overall. Moreover, some recently reported serotypes like AAV9 (Duque et al., 2009; Foust et al., 2009), AAVrh10 (Hu et al., 2010; Zhang et al., 2011), and AAV PHP capsids (Chan et al., 2017; Deverman et al., 2016) are able to cross the blood-brain barrier (BBB) upon systemic delivery. This constitutes a major advantage that opens the door for less invasive and systemic administration, illustrated perfectly by the success of Zolgensma, which used AAV9 administered intravenously to achieve remarkable therapeutic benefit in the CNS (Mendell et al., 2017; Saraiva et al., 2016). For these reasons, AAVs are the prime choice nowadays concerning gene therapy applications to the clinic and they will also be the ones used in this project. However, like every other vectors, AAVs present with disadvantages, among which

the most pressing are the following: the limited packaging capacity (~4.7 kb), which hampers the delivery of larger transgenes; the accumulation in the liver after systemic delivery, with the potential for hepatotoxicity; the existence of serum-circulating anti-AAV neutralizing antibodies in several species, including humans, which can seriously inhibit systemic gene delivery (Balakrishnan and Jayandharan, 2014; Saraiva et al., 2016). These highly specific neutralizing antibodies can exist as a result of childhood exposure or previous administration of AAV viral vectors and represents one of the most significant challenge to AAV-mediated gene therapy.

1.4.2 Gene therapeutic strategies for ALS and FTD

ALS and FTD are some of the many diseases whose patients could benefit greatly from gene therapy. Indeed, as mentioned above, disease-modifying drugs Riluzole and Edaravone have been approved for treating ALS (Abe et al., 2017; Bensimon et al., 1994; Jaiswal, 2019), but their impact on disease progression and life expectancy of patients is marginal, while FTD patients rely only on symptomatic management. Therefore, a successful gene therapy could revolutionize the treatment of these devastating disorders.

For ALS, a great many number of strategies have been developed so far in pre-clinical studies, either focusing on less specific strategies that could benefit all affected individuals, or targeting the precise causative mutations responsible for familial and some sporadic cases (Tosolini and Sleigh, 2017). A gene therapy-based global ALS treatment could theoretically be achieved by overexpressing neuroprotective genes and even by silencing a major secondary gene or disease pathway which, if successful, would benefit patients with different subtypes of the disorder. Several research groups have pursued such strategies, and a compilation of these approaches has been thoroughly reviewed (Scarrott et al., 2015). Briefly, this research avenue was explored by overexpressing neurotrophic factors such as vascular endothelial growth factor (VEGF) (Azzouz et al., 2004; Dodge et al., 2010; Krakora et al., 2013; Storkebaum et al., 2005; Wang et al., 2007), glial-derived neurotrophic factor (GDNF) (Krakora et al., 2013; Suzuki et al., 2008; Wang et al., 2002), hepatocyte growth factor (HGF) (Kadoyama et al., 2007), insulin-

like growth factor-1 (IGF-1) (Dodge et al., 2008, 2010; Kaspar et al., 2003) and granulocyte-colony stimulating factor (Henriques et al., 2011; Pitzer et al., 2008), to promote neuroprotection. Other groups have reported some interesting alternative approaches, like overexpressing anti-apoptotic genes (Azzouz et al., 2000; Yamashita et al., 2002) and downregulating intermediary pathogenic messengers like miRNAs (Koval et al., 2013), genes involved in calcium-mediated excitotoxicity (Rembach et al., 2004) and apoptosis (Locatelli et al., 2007; Scarrott et al., 2015). Nevertheless, the beneficial effects of the above-mentioned treatments are usually modest, and they are more likely to produce secondary effects, due to their non-specific nature.

With the advancement of ALS genetics and the creation of transgenic models, however, more targeted therapies started to emerge. In this regard, the greatest progress has undoubtedly been achieved in ALS with mutations in *SOD1* (SOD1-ALS), probably due to it being the first and best studied genetic origin of this disease, and the approaches have mainly focused on downregulating SOD1 using ASOs or interference RNA. The first proof-of-concept with ASOs was successful in reducing mRNA levels of SOD1, extending the lifespan and improving motor performance of $SOD1^{G93A}$ rats (Smith et al., 2006), paving the way for a phase I clinical trial in patients with SOD1-ALS to assess the toxicity of this treatment (Miller et al., 2013). In this trial, the therapy was administered intrathecally to a total of 21 participants and was deemed well tolerated, without any significant adverse effects, but it failed to reduce SOD1 protein levels in the spinal cord and CSF of patients (Miller et al., 2013). This prompted researchers to develop new, more potent ASOs that significantly extended survival and even reversed compound muscle action potential (CMAP) signatures in $SOD1^{G93A}$ ALS rodent models (McCampbell et al., 2018). Excitingly, these next-generation ASOs are currently being tested in human clinical trials and have already been shown to reduce SOD1 levels in the CSF by up to 33 percentage points, following regular intrathecal delivery into patients with SOD1 mutations (Miller et al., 2020).

As for interference RNA approaches, these have been delivered by viral vectors, first using lentiviruses to express a short-hairpin RNA (shRNA) and significantly improving symptoms and survival of $SOD1^{G93A}$ transgenic mice (Ralph et al., 2005; Raoul et al., 2005). However, further developments of the safer AAV viral vectors, in particular the neurotropic serotypes AAV9 and AAVrh10, prompted important proof-of-concept studies in which delivery of shRNAs (Foust et al., 2013; Iannitti et al., 2018; Thomsen et al., 2019) or artificial miRNA (Borel et al., 2016; Keeler et al., 2020; Stoica et al., 2016) delayed disease progression and

prolonged survival in SOD1 mouse models. Importantly, experiments in NHPs validated the clinical potential of these approaches (Borel et al., 2016; Foust et al., 2013; Thomsen et al., 2019), leading to the recent treatment of two SOD1-ALS patients with AAVrh10 delivering an artificial SOD1-targeting miRNA, considered safer than shRNAs (Mueller et al., 2020). Delivered through a single intrathecal injection, the treatment was considered relatively safe for use in humans, although immunosuppression was deemed necessary to prevent an adverse inflammatory response (Mueller et al., 2020). Despite decreased levels of SOD1 detected in the spinal cord of one patient, no assertions can be made about the clinical benefit of this therapy, which will need to be determined by a larger study. Other approaches have been proposed to treat SOD1-ALS using gene editing, which will be discussed below.

Despite being discovered almost two decades after the first *SOD1* mutations, the hexanucleotide expansion in the C9orf72 gene is the most frequent genetic origin of ALS and FTD, which prompted scientists to develop new targeted gene therapy approaches at increased speed. Of these, ASOs binding to the C9orf72 gene are by far the most advanced, having been shown to consistently inhibit accumulation of expanded transcripts and, consequently, decrease RNA foci in iPSC-derived neurons (Aoki et al., 2017; Donnelly et al., 2013; Lagier-Tourenne et al., 2013; Sareen et al., 2013). In addition to this, ASOs were successful in downregulating C9orf72 *in vivo*, decreasing G4C2 by-products and, in some cases, rescuing disease-associated pathways and behavioural deficits in *D. melanogaster* (Zhang et al., 2015) or mouse models (Jiang et al., 2016; O'Rourke et al., 2015). These successes motivated the design of a human clinical trial which is now ongoing (NCT03626012) to assess the safety, tolerability, and pharmacokinetics of the ASO therapy. As an alternative to ASO approaches, however, AAV-mediated delivery of artificial miRNAs has also proven successful in reducing RNA Foci and poly-(GP) DPRs in primary neurons (Peters et al., 2015), and reducing Foci in iPSC-derived neurons and in a transgenic C9orf72 mouse model (Martier et al., 2019a, 2019b). In addition to these strategies, gene editing has been explored as a potential therapeutic tool for C9orf72-ALS/FTD and will be discussed in more detail below.

For other genetic subtypes of ALS, gene therapy approaches are considerably less advanced. A young patient with a fast-progressing form of ALS caused by a FUS mutation has been given a tailored ASO treatment (Arnold, 2019), perhaps paving the way for other personalised FUS-targeting strategies. Also, in an exciting study, Becker and colleagues (Becker et al., 2017) used ASOs targeting a gene modifier of TDP-43, *ATXN2*, to treat TDP-43 transgenic mice, observing remarkable improvements in motor performance and survival of these animals. This

report raises hopes for a future gene therapy treatment for ALS with *TARDBP* mutations and, considering how TDP-43 aggregation is a feature in ~97% of cases, it might even have broader applicability to other fALS and sALS cases (Becker et al., 2017).

Due to the significant genetic overlap between ALS and FTD, the gene therapy approaches mentioned above that target the *C9orf72* gene, *TARDBP*, and *FUS* would also be applicable for patients suffering from FTD. In addition to these, AAV1-mediated gene replacement of *GRN* was successful in increasing progranulin in the brain and reversing FTD-like behavioural aspects in $Grn^{+/-}$ mice (Arrant et al., 2017). Excitingly, a phase 1-2 clinical trial is set to start recruiting this year to test an AAV9-based gene therapy to increase progranulin levels in the brain, following a one-time intra-cisternal injection into the CSF (NCT04408625).

In summary, the significant progress and milestones described above raise hopes that patients afflicted with these devastating and fatal diseases might soon be offered more efficient therapies that could slow their disease progression, extend survival and improve their quality of life. In addition to the strategies mentioned above, however, new possibilities are arising fast with the advent of gene-editing technologies such as zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENS), meganucleases, and the clustered regularly interspaced short palindromic repeats (CRISPR)-Cas. Indeed, these technologies are now significantly complementing the gene therapy toolkit by offering unprecedented access to directly correct disease-causing genomic mutations (Maeder and Gersbach, 2016). Due to its simplicity and versatility, though, the CRISPR-Cas system has particularly stood out among the other editing techniques and will be discussed in more detail bellow, focusing mainly on the highly successful CRISPR-Cas9 tool.

1.4.3 CRISPR-Cas9 therapeutic gene editing

CRISPR is a bacterial adaptive “immune system” against bacteriophages in which chunks of the viral genome, termed spacers, are introduced in a specific bacterial *locus* - the CRISPR-Cas array - allowing to “memorize” the invading genome for future infections. In the class 2 system, of which we will focus on the type II CRISPR-Cas9, this array is composed by the

following elements: the CRISPR array, where the spacers can be inserted in several different sites, all of which intercalated by repeat sequences; the CRISPR-associated (Cas) operon, located upstream, which codifies to the effector Cas nuclease responsible for introducing breaks in the invading DNA, Cas9, as well as Cas1, Cas2 and Csn2 proteins, important for the spacer acquisition step; and the tracrRNA, upstream of the Cas operon, which is crucial for maturation of the precursor CRISPR RNA (pre-crRNA) (Doudna and Charpentier, 2014; Hsu et al., 2014). Following the transcription of the CRISPR array into pre-crRNA, the tracrRNA forms a duplex with the complementary repeat sequence of the precursor molecule, which is stabilized by Cas9. Then, upon a subsequent infection by the same virus, the Cas9 protein is guided by the tracrRNA:crRNA duplex into the complementary viral sequence and introduces a precise blunt double-stranded DNA (dsDNA) break, disrupting in this way the viral infection. The dsDNA break is exerted by the two catalytic domains of the Cas9 nuclease – RuvC and HNH – and is dependent upon the presence of a protospacer adjacent motif (PAM) contiguous to the targeted DNA sequence (Hille and Charpentier, 2016; Mali et al., 2013a). The PAM is comprised of a nucleotide string that varies according to the Cas9 orthologue, with the most common ones being 'NGG' for *S. pyogenes* Cas9 (SpCas9); 'NNGRR(T)' for *S. aureus*, 'NNAGAAW' for *S. thermophilus*, NNNNGATT for *Neisseria meningitides*, among others.

Despite the existence of alternative types of bacterial adaptive immunity, including the class I and class III CRISPR systems, the simplicity of CRISPR-Cas9 makes it more appealing to employ in gene editing projects. A major step forward in this direction was taken by engineering a chimera, termed single-guide RNA (sgRNA), with two major components: a double stranded RNA molecule composed of the 3' end crRNA linked to the 5' end of the tracrRNA by a synthetic stem loop, called the sgRNA scaffold; and the 5' end of the crRNA containing a variable ~20 nucleotide string, complementary to the targeted DNA sequence (Jinek et al., 2012). This 'upgrade' bypasses the need for a dual-gRNA system and opened the door to turn CRISPR-Cas9 into an effective tool capable of inserting cuts in the DNA with remarkable precision *in vitro* (Jinek et al., 2012), and one year later in mammalian cells (Cong et al., 2013; Mali et al., 2013b), revolutionizing the field of genome editing. The potential applications are exceptionally diverse, encompassing many fields of fundamental biology, but also agricultural research, disease modelling and medicine (Hsu et al., 2014). Indeed, the introduction of plasmid DNA encoding for Cas9 and a U6 promoter-driven sgRNA in almost any kind of cell, whether it is by means of viral or non-viral transfection, is sufficient to introduce a dsDNA break in a precise region of the host genome, which the cell then tries to

repair by one of two methods: non-homologous end joining (NHEJ) or homology directed repair (HDR). The first is error-prone and often results in insertions/ deletions (indels) and, if the cutting takes place in an exon, the introduction of stop codons that disrupt the expression of the gene (Ran et al., 2013a). Conversely, HDR is highly reliable and accurate. By co-delivering a homology repair DNA template of choice with the Cas9-sgRNA complex, it becomes theoretically possible to introduce any type of nucleotide sequence (up to 1 kb at least) in a desired genomic location (Ran et al., 2013a), although the efficiency of this process is usually low in most cell types (Figure 1-5).

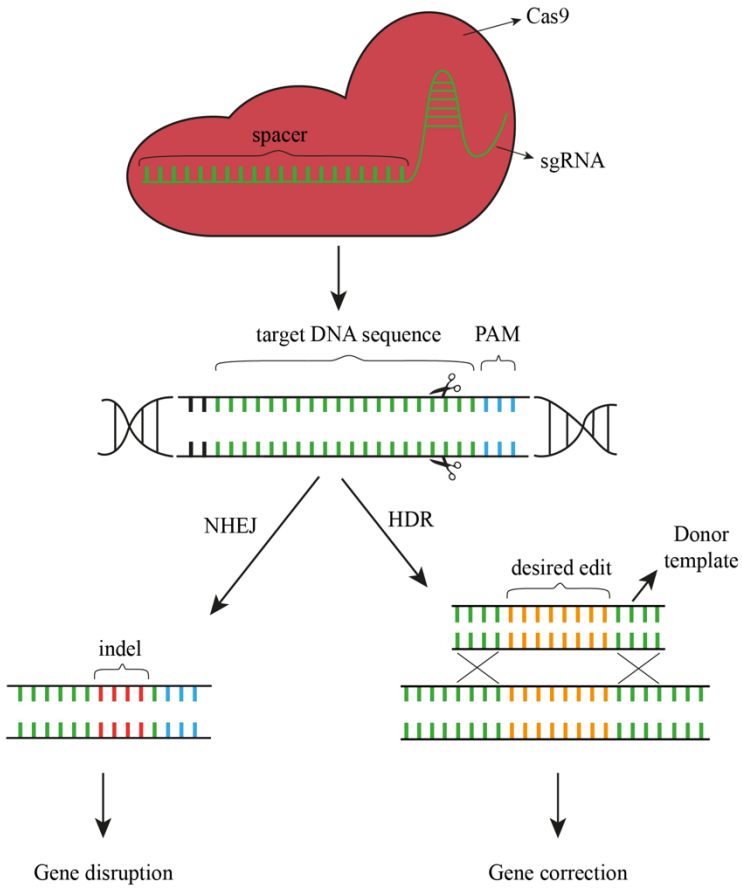


Figure 1-5. CRISPR-Cas9-mediated gene editing. A stable complex formed by Cas9 and a sgRNA is directed to a targeted genomic region with sequence complementarity to the spacer and a PAM sequence. Presence of the PAM is necessary for the edit and its sequence varies with the Cas9 orthologue, the most common one being 'NGG' for SpCas9. The two catalytic domains of Cas9 then introduce a dsDNA break at 3-4 bp from the PAM, represented by black scissors, which the cell repairs

by either NHEJ or HDR. Being error-prone, NHEJ results in random small indels forming in the cut site which, if directed to an exon, can introduce premature stop codons and disrupt the expression of a specific gene. If, however, a donor template with overlapping arms is supplied, the highly reliable but low-efficiency HDR can happen, allowing to introduce almost any desired edit, from single base changes to indels.

Despite its versatility, efficiency and simplicity, a major concern arose soon after the first CRISPR-Cas9 reports: the off-target effects. Indeed, three independent studies confirmed that, even with sgRNAs possessing up to 5 mismatches with the targeted sequence, Cas9 is still able to introduce dsDNA breaks in human cells (Fu et al., 2013; Hsu et al., 2013; Pattanayak et al., 2013). Although the efficiency of this off-target activity is considerably reduced when comparing with the on-target, it became evident that much refinement and optimizations still had to occur before considering the use of this technology for commercial and biomedical applications. Owing to the astonishingly fast pace of the field, however, remarkably elegant and creative solutions have already been put forward to increase the specificity of this tool, or to bypass some of its potential off-target deleterious effects. Of these upgrades, we will mention the most impactful ones, starting with the creation of search algorithms, most of them as freely accessible software, that allow to select sgRNAs with the least computationally predicted off-target cut sites (Bae et al., 2014; Haeussler et al., 2016; Montague et al., 2014; Ran et al., 2013a). Several approaches focused instead on creating high-fidelity variants of the Cas9-sgRNA complex. This was achieved by either modifying specific aa residues of the Cas nuclease (Chen et al., 2017; Kleinstiver et al., 2016; Slaymaker et al., 2016), leading to better discrimination between on-target and off-target activity, or by introducing chemical modifications in the sgRNAs that significantly reduce off-target binding (Ryan et al., 2018). In addition to these improvements, the relentless expansion of the CRISPR toolkit recently allowed new editing applications that, although still generating off-target effects, do not directly rely on blunt dsDNA cutting, thereby minimizing the danger of unwanted gene disruptions and chromosomal rearrangements. These upgrades were achieved by using Cas9 variants with engineered mutations in the catalytic domains of the nuclease. One of these variants was created by mutating both catalytic domains, forming a catalytically-dead Cas9 (dCas9) which still retains its ability to bind DNA in a sgRNA-mediated programmed manner (Jinek et al., 2012). This dCas9 was then used as a core component to engineer new molecular

tools in recent years, including CRISPR interference (CRISPRi). This interesting technique complements interference RNA by using dCas9 alone or fused to effector domains that either repress or activate transcription, becoming a powerful tool to regulate gene expression (Figure 1-6 A) (Gilbert et al., 2013; Qi et al., 2013). In addition, several authors demonstrated how the CRISPR-Cas9 system can also be repurposed to bind, track and cut RNA in a programmed manner, either by the use of PAM-presenting oligonucleotides (PAMmers) or by fusing dCas9 to an RNA endonuclease domain (Figure 1-6 B) (Batra et al., 2017; Nelles et al., 2016; O'Connell et al., 2014). Another interesting approach, however, consisted in mutating only one of the catalytic domains of Cas9 (RuvC) which, instead of dsDNA breaks, is only capable of introducing single-stranded DNA breaks (ssDNA breaks) in the host genome, easily repaired by the cell machinery. By combining this mutated Cas9 with a pair of sgRNAs targeted at both the sense and antisense strands of the DNA, in a method named 'double nickase', Ran and colleagues (Ran et al., 2013b) were able to artificially introduce dsDNA breaks in the desired location while decreasing the off-target cutting by a factor of several hundred-fold (Figure 1-6 C). Also, as an alternative application to the ones mentioned above, a new technique first developed in 2016 by Komor and colleagues revolutionized the field once again by allowing programmable base changes from a C-G pairing into T-A, without effecting dsDNA cuts (Komor et al., 2016). This creative solution, named 'base editing', was conceived by fusing a RuvC-mutated Cas9 nickase with a cytosine deaminase enzyme that converts cytosine to Uracil, which has the base-pairing properties of thymine (T). The complex was then also fused with a uracil glycosylase inhibitor (UGI) to prevent the reversing of this reaction and, by nicking the non-edited DNA strand, this cytosine base editor (CBE) favors in this way the correction of the U-G mismatch into the desired T-A pairing (Komor et al., 2016). The base editing toolkit was eventually upgraded to allow the programmed conversion of A-T pairing into G-C, by using a similar principle but instead fusing the catalytically impaired Cas9 to an artificially evolved adenine deaminase, creating adenine base editors (ABEs) (Figure 1-6 D) (Gaudelli et al., 2017). Finally, a recently developed technique, termed prime editing, promises to take the gene editing field even further in the quest to correct almost all pathogenic mutations (Anzalone et al., 2019). Indeed, this innovative tool is capable of introducing insertions, deletions, and any base-to-base conversions in a programmable manner, without requiring dsDNA breaks or donor DNA templates. To achieve this, a nickase Cas9 fused to a reverse transcriptase (RT) domain is guided to the targeted sequence by a prime editing guide RNA (pegRNA), where it synthesizes a small DNA fragment from a template, located within the pegRNA, containing the desired edit (Figure 1-6 E) (Anzalone et al., 2019).

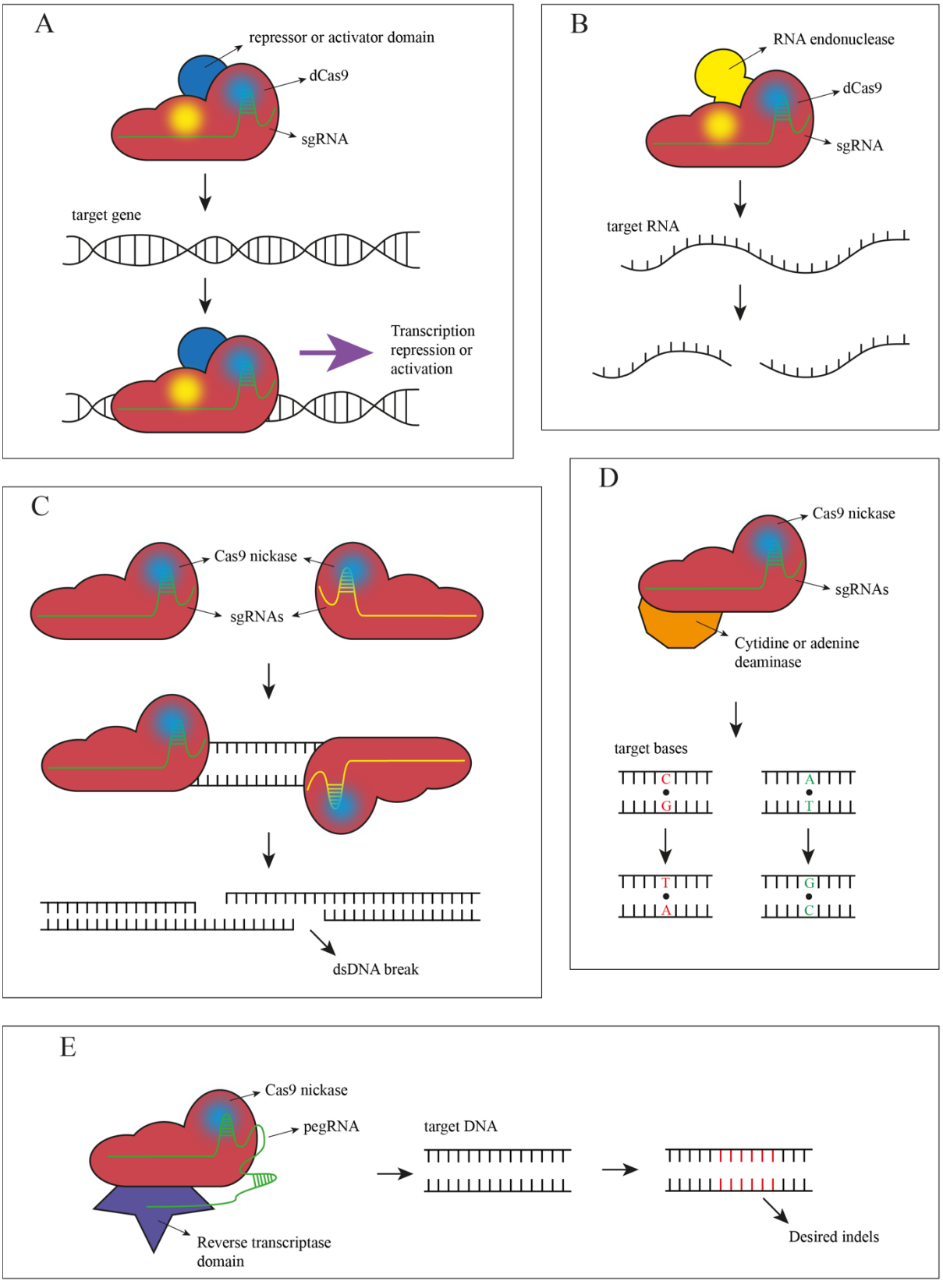

A
repressor or activator domain
dCas9
sgRNA
target gene
Transcription
repression or
activation
B
RNA endonuclease
dCas9
sgRNA
target RNA
C
Cas9 nickase
sgRNAs
dsDNA break
D
Cas9 nickase
sgRNAs
Cytidine or adenine
deaminase
target bases
C
G
A
T
T
A
G
C
E
Cas9 nickase
pegRNA
Reverse transcriptase
domain
target DNA
Desired indels

Figure 1-6. CRISPR-Cas9-based alternative editing tools with increased specificity. Alternative CRISPR-Cas9 gene editing tools have been developed based on Cas9 variants with mutations in its catalytic domains, preventing blunt dsDNA cuts and reducing safety and ethical issues of translational applications. (A) Catalytically dead Cas9 (dCas9), created by mutating both catalytic domains, can be used alone or fused to a transcription activator/repressor domain to bind target DNA in a programmed manner. Although not strictly an editor in itself, this tool allows for controlled regulation of gene expression by activating or repressing transcription. (B) dCas9 can also be fused to an RNA endonuclease, being able to cleave target RNA molecules. (C) The 'double nickase' technique uses a catalytically impaired Cas9, or nickase Cas9, created by mutating only one of its catalytic domains, resulting in the nuclease introducing only ssDNA breaks, which are easily repaired by the cell. Combining the nickase Cas9 with a pair of sgRNAs binding close to each other can artificially create a double-strand break with long overhangs on each cleaved end instead of a blunt cut. This allows for gene disruption and HDR-based editing with remarkably increased specificity. (D) Base editing can be achieved by fusing nickase Cas9 with a cytosine or an adenine deaminase, allowing to directly convert target bases without dsDNA breaks. (E) Prime editing also uses nickase Cas9, but instead fused to a reverse transcriptase domain. This complex is guided by the pegRNA to the target sequence, where it allows for the introduction of any desired small indel without dsDNA cut, by reverse transcribing a provided template.

As well as many other scientific fields, the CRISPR-Cas9 system holds great promise for gene therapy, opening the door to new approaches such as correction of genetic defects with remarkable precision, instead of just overexpression or silencing paradigms (McMahon and Cleveland, 2016; Xiao-Jie et al., 2015; Xue et al., 2016). The first attempts at editing disease-causing mutations were made *in vitro*, particularly in the zygote of a mouse harboring a dominant deletion in the exon 3 of the *CRYGC* gene, which causes cataracts. Through HDR, the mutant allele was successfully repaired with minimal off-target effects, and the zygote was capable of originating viable, gene-corrected progeny (Wu et al., 2013). Following this first proof-of-principle, some impressive works were also reported in other diseases, of which we will mention a few examples: the correction of the F508 deletion in the *CFTR* gene, responsible for cystic fibrosis, in intestinal stem cells from patients (Schwank et al., 2013); the introduction of the natural occurring CCR5Δ32 mutation in iPSCs from healthy individuals, conferring resistance to the HIV virus after differentiation into monocytes/ macrophages (Ye et al., 2014); the multi exon-skipping of a large region of the *DMD* gene (exons 45-55) in myoblasts from

people with Duchenne muscular dystrophy, with the consequent restoration of dystrophin expression (Ousterout et al., 2015); the allele-specific disruption of mutant *Huntingtin* in various cell lines from human patients (Shin et al., 2016); the germline correction of the nonsense mutation in the *DMD* gene of mdx mice, creating viable progeny with significant recovery of dystrophy expression and mitigation of dystrophic muscle phenotype (Long et al., 2014). Despite these early exciting advances *in vitro*, however, the utopic concept of therapeutic gene editing has long been the ability to edit the genome *in vivo*, although the ideal delivery system has continuously eluded scientists, as discussed above (section 1.4.1). Hao and colleagues took advantage of the hydrodynamic tail vein injection of naked DNA, a procedure known for its ability to transfect liver cells in mice, to establish the first demonstration of an *in vivo* gene editing experiment in adult animals (Yin et al., 2014b). The authors of this study injected Cas9 with a sgRNA targeted to the *FAH* gene in a mouse model of a fatal liver disease, hereditary tyrosinemia type I, along with a repair template to allow for homologous recombination. This experiment was successful in inducing FAH^+ hepatocytes at approximately 30 days after treatment, in addition to reducing liver damage and attenuating the weight loss associated with the disorder. Despite this elegant approach, the transduction of other organs and tissues demands the use of a suitable vector, among which AAVs stand out as both safe and efficient for human gene transfer. Their packaging size limit, however, renders these viruses impractical for delivering both Cas9 and a sgRNA in a single construct, for which reason Swiech and co-workers (Swiech et al., 2014) relied on a double-injection paradigm, with the sgRNAs being delivered in a separate batch of viruses. In 2015, the same research group published a significant breakthrough by characterizing a Cas9 orthologue from a different species, *Staphylococcus aureus* (herein, SaCas9), with equivalent cutting efficiency and a smaller size (~3.2 kb) than the standard *S. pyogenes* orthologue (Ran et al., 2015). The authors then validated the use of this SaCas9 *in vivo* by targeting the murine *Apob* gene in the liver, upon intravenous injection of C57BL/6 mice with recombinant AAV serotype 8. Finally, three independent groups went one step further and reported a successful exon skipping strategy for Duchenne muscular dystrophy, using different AAV serotypes (AAV8 and AAV9) with either SpCas9 or SaCas9 delivered through various systemic routes (Long et al., 2016; Nelson et al., 2016; Tabebordbar et al., 2016). Indeed, by delivering sgRNAs specific to the flanking regions of the mutated exon 23 of mdx mice, the authors were able to excise that exon and still obtain an apparently functional version of the dystrophin protein.

Countless other therapeutic proof-of-concept studies have been done using CRISPR-Cas9, and have been reviewed elsewhere (Karimian et al., 2019; Xiao-Jie et al., 2015), but we will briefly mention the ones relevant to ALS and FTD, namely the proof-of-concept therapies for SOD1-ALS and C9orf72-ALS/FTD. For SOD1-ALS, the first attempts used intravenous (Gaj et al., 2017) or intracerebroventricular (ICV) (Duan et al., 2020) injection of AAVs delivering SaCas9 to create disruptive indels in mutant SOD1, inhibiting its expression, resulting in delayed disease onset and increased survival of $SOD1^{G93A}$ mice. Using the same mouse model, Lim and colleagues (Lim et al., 2020) recently employed an alternative base editing approach, using a CBE split across two AAV viral vectors. Upon intrathecal delivery, this dual-vector approach was successful in introducing a nonsense-coding substitution into the mutant SOD1 gene, improving survival and slowing down disease progression (Lim et al., 2020). For C9orf72-ALS/FTD, CRISPR-based therapies were first explored by using RNA-targeting Cas9 to cleave G4C2 expanded transcripts in a cell line, promoting a reduction of RNA Foci accumulation (Batra et al., 2017). Finally, a different strategy was also reported in which the direct targeting of G4C2 DNA with dCas9 can block the transcription and, consequently, the RAN translation of the hexanucleotide expansion, and was shown to decrease poly-(GP) DPRs *in vitro* (Pinto et al., 2017). However, considering that repetitive G4C2 DNA or RNA sequences abundantly exist in the human genome, the two approaches mentioned above would probably result in considerable off-target genomic and transcriptomic effects by directly targeting Cas9 to these sequences. Therefore, additional innovative ideas and strategies are urgently needed for treat C9orf72-ALS/FTD.

1.5 Introduction to PhD project: aims and objectives

As a consequence of being heterogeneous, polygenic syndromes, difficult to diagnose and with a high variability of symptoms and outcomes, patients diagnosed with ALS and/or FTD are still offered a dramatically small and inefficient list of therapeutic options. Indeed, as mentioned above, the treatment of FTD still relies on symptoms' management while Riluzole and Edaravone remain the only disease-modifying treatments currently approved for ALS, with very modest effects on disease progression and survival. However, the last few years have seen

remarkable advances in the investigation of these devastating disorders, much owing to increased media attention and public awareness, as well as increasingly powerful technologies such as next-generation sequencing. As a result, the discovery of the hexanucleotide repeat expansion in the C9orf72 as the major genetic origin for both ALS and FTD (DeJesus-Hernandez et al., 2011; Renton et al., 2011), and the ongoing venture to understand its pathophysiology, disclosed a major part of the mystery link between both diseases. Perhaps even more importantly, however, these discoveries have suddenly opened a new horizon of possibilities for more targeted therapies, which are in urgent need, considering the failure of both drug-based and gene-based therapies aimed at a global treatment.

It was in this context that we proposed to develop and test two new experimental treatments, both consisting of a highly targeted gene editing strategy, using the CRISPR-Cas9 tool. Our first approach focuses on expressing SaCas9 through an AAV delivery viral vector together with a pair of sgRNAs that cut in the flanking regions of the hexanucleotide expansion, excising the repeats and allowing the re-ligation of the gene (C901, Figure 1-7). Excision of the repeats was successfully done by others to create control isogenic lines (Ababneh et al., 2020; Abo-Rady et al., 2020; Dafinca et al., 2020; Selvaraj et al., 2018), but its use in the context of a therapeutic application has, to the best of our knowledge, never been explored. If successful, this approach has the potential to tackle all the C9orf72-related pathological mechanisms mentioned previously, including the haploinsufficiency of C9orf72 protein, RNA-mediated toxicity, and RAN translation resulting in the accumulation of DPRs.

Our second strategy focuses mainly on the presence of DPRs as a pathological pathway. Indeed, the second therapeutic construct contains two sgRNAs designed to introduce a deletion in the sequence immediately upstream of the expansion, eliminating a Kozak consensus with an alternative start codon necessary for the occurrence of RAN translation (C904, Figure 1-7). Indeed, it has been demonstrated by others that expressing G4C2 repeats with mutated versions of this alternative codon significantly reduces DPR formation *in vitro* (Cheng et al., 2018; Green et al., 2017; Tabet et al., 2018). We hypothesized that transduction with our viruses will eliminate this highly regulatory region and might reduce the accumulation of C9orf72-derived DPRs. If successful, this second approach should theoretically act only on the products of RAN translation, as opposed to the first strategy. However, considering the highly pathogenic nature of these dipeptides, a significant reduction of their expression could be enough to alleviate the neurodegenerative phenotype of C9orf72-ALS/FTD. In addition, our second strategy involves deleting a considerably smaller sequence of DNA, approximately 90 base pairs (bp) in contrast

with the thousands of the repeat expansion, which might significantly increase the efficiency of the editing.

Therefore, the goals for this project can be structured into four main objectives/aims: a) design and prepare single or dual pAAV constructs capable of expressing SaCas9 and any pair of U6-driven sgRNAs to create targeted deletions in any region of the genome. Design and clone different sgRNA sequences that target the human C9orf72 gene according to our two therapeutic strategies and validate these approaches in a human cell line with 3xG4C2 repeats; b) produce AAV9 viruses and test whether these viral vectors can edit patient-derived expanded C9orf72 in primary neurons, and whether this editing alleviates C9orf72-specific pathological hallmarks; c) evaluate the capability of our two experimental therapies to edit the human C9orf72 gene *in vivo* and whether they ameliorate the phenotype of a BAC-transgenic mouse model with C9orf72-linked neurodegeneration and decreased survival; d) introduce modifications to the strategies used, such as AAV capsid with improved CNS tropism and stronger promoter, to optimize the delivery and expression of the therapeutic cassettes and test whether these modifications maximize the efficiency of our proof-of-concept.

The highly innovative nature of this project has the potential to widen the spectrum of CRISPR-Cas9 applications in molecular medicine, consisting of two new approaches to the treatment of C9orf72-ALS/FTD, with potential applicability to other nucleotide repeat expansion disorders.

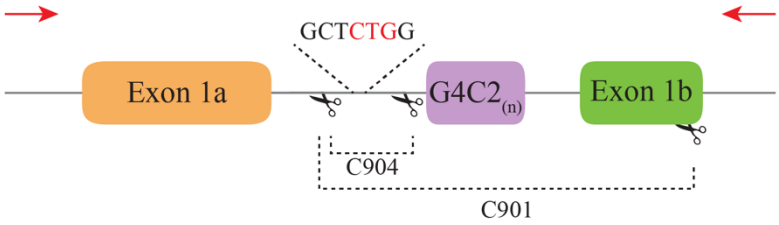


Figure 1-7. Schematic representation of the two therapeutic strategies designed and detailed in this project. Schematic outline of our two therapeutic strategies, containing the initial portion of the C9orf72 gene, with the hexanucleotide repeat region between the non-coding exons 1a and 1b and the scissors representing the cut sites for the different sgRNAs tested. The C901 treatment cuts on either side of the expansion to mutate the flanking sequences and facilitate excision of the G4C2

repeats. C904 treatment disrupts the sequence directly upstream of the repeats, promoting the excision of a ~90 bp fragment that contains a strong Kozak consensus sequence. Red arrows represent C9_OUT primers.

2. Materials and methods

2.1 Materials

Throughout this thesis, the term 'vector' refers to the plasmid DNA or it can be used in specific sections to ambiguously designate both the plasmid DNA and the viral particle that packages it. When used exclusively in the context of the viral vector, the term 'viral' is included in addition to 'vector'.

2.1.1 Plasmid DNA

Table 2.1. Plasmid DNA. List of plasmid DNAs used in this project.

Plasmid	Obtained from:
pX601	pX601-AAV-CMV::NLS-SaCas9-NLS-3xHA-bGHpA;U6::BsaI-sgRNA was a gift from Feng Zhang (Addgene plasmid #61591; http://n2t.net/addgene:61591 ; RRID:Addgene_61591) (Ran et al., 2015). Sequence available in Addgene.
LentiCRISPR V2	lentiCRISPR v2 was a gift from Feng Zhang (Addgene plasmid # 52961 ; http://n2t.net/addgene:52961 ; RRID:Addgene_52961) (Sanjana et al., 2014). Sequence available in Addgene.
pX601::dual sgRNA cassette	Cloned in-house (see 2.2.1.7 Cloning of the pAAV-CRISPR-Cas9 single-vector). Sequence available in **appendix 1**.
pX601::EFS1a::dual sgRNA cassette	Cloned in-house (see 2.2.1.7 Cloning of the pAAV-CRISPR-Cas9 single-vector). Sequence available in **appendix 1**.
pAAV-CRISPR-Cas9 single-vector	Cloned in-house (see 2.2.1.7 Cloning of the pAAV-CRISPR-Cas9 single-vector). Sequence available in **appendix 1**.
pC901	Cloned in-house (see 2.2.1.2 Cloning of sgRNAs). Sequence available in **appendix 1**.

pC902	Cloned in-house (see 2.2.1.2 Cloning of sgRNAs). Sequence available in **appendix 1**.
pC904	Cloned in-house (see 2.2.1.2 Cloning of sgRNAs). Sequence available in **appendix 1**.
pHelper	Stratagene; Stockport, UK
AAV9 Rep-Cap	Kindly provided by James Wilson, University of Pennsylvania
AAV PHP.eB Rep-Cap	pUCmini-iCAP-PHP.eB was a gift from Viviana Gradinaru (Addgene plasmid # 103005 ; http://n2t.net/addgene:103005 ; RRID:Addgene_103005) (Chan et al., 2017). Sequence available in Addgene.

2.1.2 Viral vectors

Table 2.2. Viral vectors. List of viral vectors used in this study.

Virus	Produced:
AAV9 C901	In-house (see 2.2.1.11 AAV production)
AAV9 C901 (2)	In-house (see 2.2.1.11 AAV production)
AAV9 C902	In-house (see 2.2.1.11 AAV production)
AAV C902 (2)	In-house (see 2.2.1.11 AAV production)
AAV9 C904	In-house (see 2.2.1.11 AAV production)
AAV9 C904-CHOP	Children's Hospital of Philadelphia clinical vector core (CHOP)
AAV PHP.eB CMV-GFP	In-house (see 2.2.1.11 AAV production)

AAV PHP.eB pX601 Ctrl3	In-house (see 2.2.1.11 AAV production)
AAV PHP.eB pX601 T4I	In-house (see 2.2.1.11 AAV production)
AAV PHP.eB pX601 T4D	In-house (see 2.2.1.11 AAV production)
AAV PHP.eB pX601 T5B	In-house (see 2.2.1.11 AAV production)

2.1.3 gBlocks

Table 2.3. gBlocks sequences. The sequences of the gBlocks used in this project, as they were ordered from IDT.

Name	Sequence (5'-3')
L-gBlock	CAGTCAGTCGTCTCACACCAGGGTACCTGAGGGCCTATTTCCCAT GATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGAT AATTGGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAA ATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTT TAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTT GAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGAC GAAACACCGGAGACCACGGCAGGTCTCAGTTTTAGTACTCTGGA AACAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTATCTCGTC AACTTGTTGGCGAGATTTTTGCAGAGACGATCTAGCA
R-gBlock	GATACGTACGTCTCCTTGCCTGAGGGCCTATTTCCCATGATTCCTT CATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGGA ATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTG ACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATT ATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTA TTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACAC CGGGTCTTCGGATCCATGCGAAGACCTGTTTTAGTACTCTGGAAA CAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTATCTCGTCAA

	CTTGTTGGCGAGATTTTTGCGGCCGCCGTGGTTTTGAGACGATTC AGATG

2.1.4 DNA oligonucleotides

Table 2.4. List of DNA oligonucleotides. The sequences of PCR primers, sequencing primers and sPoly(A) oligonucleotides as they were ordered from Sigma.

Name	Sequence (5'-3')
U6F	ACTATCATATGCTTACCGTAAC
SV4F	GAAGAGAATAGCAGGCATGCT
T4BF	GAAACACCGCGGGGTCTAGCAA
T4IF	GAAACACCGCTTGCTCTCACA
Ctrl1F	GAAACACCGATCAGTTCGACG
C9_OUT_7_Fwd	CGTCATCGCACATAGAAAACAGA
C9_OUT_7_Rev	ACCAGTCGCTAGAGGCGAA
C9_OUT_2_Fwd	TTCGCTAGCCTCGTGAGAAAA
C9_OUT_2_Rev	CCAGTCGCTAGAGGCGAAAG
SV3F	GAAACACCGGAGACCACGGCA
SV3R	GCCGTGGTCTCCGGTGTTT
SV5F	GATCCATGCGAAGACCTGTT
EFS1a_Fwd	AGACGTCGTCTAGAGGCTCCGGTGCCCGTCAG

EFS1a_Rev	CTGAGCATCCATGGCCTGTGTTCTGGCGGCAAAC
Prm_Rev	GCCCAGGATGTAGTTCCGCTT
sPoly(A)_sense	AATTCAATAAAAGATCTTTATTTTCATTAGATCTGTGTGT TGGTTTTTTGTGTGCGGTAC
sPoly(A)_antisense	GATACGTACGTCTCCTTGCCTGAGGGCCTATTTCCCATGA TTCCTTCATATTTGCATAT
sPAF	GGATCCTACCCATACGAT
M13_Fwd	GTAAAACGACGGCCAG
C9_INT_1_Fwd	GGTGTGGGTTTAGGAGGTGT
C9_INT_1_Rev	CGACTCCTGAGTTCCAGAGC
C9_EXT_3_Fwd	TGCGGTTGCGGTGCC
C9_EXT_3_Rev	CTCCTGGGAAAGTGCAGGAC
C904_INT_14_Fwd	GCTAGCCTCGTGAGAAAACG
C904_INT_14_Rev	TAGCGCGCGACTCCTG
Cas9_Fwd	ATCGACTACGAGACACGGGA
Cas9_Rev	CTGGATTCTATGCCGCCTCC
21238	CTGTCCCTGTATGCCTCTGG
21239	AGATGGAGAAAGGACTAGGCTACA
35766	TCGAAATGCAGAGAGTGGTG
35767	CTTCCTTTCCGGATTATATGTG
GE_Cut_Fwd	CTCACAGTACTCTGGGTGAGT
GE_Cut_Rev	CAAGGAAGAGGCCAGATCCC
GE_Ctrl_Fwd	ATCCAGCAGCCTCCCCTAT

GE_Ctrl_Rev	CCAAAAGAGAAGCAACCGGG
MRX-F	FAM-TGTAAAACGACGGCCAGTCAAGGAGGGAAACAACCGCAGCC
MRX-M13R	CAGGAAACAGCTATGACC
MRX-R1	CAGGAAACAGCTATGACCGGGCCCGCCCCGACCACGCCCCGGCCCCGGCCCCGG

2.1.5 Single guide-RNAs

Table 2.5. Single guide-RNA sequences. The sequences of the different sgRNAs used in this project are listed bellow. 'NNGRRT' was the PAM sequence used to design all the sgRNAs.

Name	Sequence (5'-3')-'NNGRRT'
T4I/T4U	CTTGCTCTCACAGTACTCGCT
T4B	CGGGGTCTAGCAAGAGCAGGT
T5B	GGCGCAGGCGGTGGCGAGTGG
T5I	GCATCCTGGCGGGTGGCTGTT
T7I	GAACCCCAAACAGCCACCCGC
T4D	CTGTAGCAAGCTCTGGAACTC
Ctrl1	GATCAGTTCGACGCGATACGGA
Ctrl2	GCGATCCGAATAGCTCGATGC

2.2 Methods

2.2.1 *In vitro* experimental methods

2.2.1.1 Design of sgRNAs targeting C9orf72

The sequence of the human C9orf72 gene was retrieved from the ENSEMBL database (Yates et al., 2016) and used as template for the software Breaking-Cas (http://bioinfogp.cnb.csic.es/tools/ breakingcas/) (Oliveros et al., 2016). Although SaCas9 still has appreciable cutting with 'NGRRN' PAM sequences, its nuclease activity is maximal with 'NGRRT' as PAM (Ran et al., 2015), for which reason this was the sequence selected by us. One of our two therapeutic approaches focuses on excising the hexanucleotide expansion, allowing the re-ligation of the gene. Considering this, we designed two sgRNAs cutting upstream of the expansion (T4B and T4I) and 3 cutting downstream (T5B, T5I and T7I), chosen based on minimal predicted off-target effects and proximity to the G4C2 sequence (Table 2.5). For our second therapeutic option, we designed one sgRNA cutting upstream of the targeted 'CUG' start codon (same sequence as T4I sgRNA, but denominated T4U when mentioning this approach) and a second sgRNA cutting downstream, only a few bp away from the expansion (T4D) (Table 2.5).

The sgRNAs were ordered as single-stranded DNA oligonucleotides and were posteriorly annealed and cloned into vectors containing the appropriate cloning cassettes.

2.2.1.2 Cloning of sgRNAs

The AAV plasmid expressing SaCas9 and one sgRNA cassette (pX601) was a gift from Feng Zhang (Ran et al., 2015) and the cloning of sgRNAs into this vector was adapted from a

previously published protocol (Ran et al., 2013a), with some changes. Briefly, the ssDNA oligonucleotides designed *in silico* were annealed and phosphorylated in a single step, using a thermocycler and the T4 Polynucleotide Kinase (T4 PNK, NEB). The thermocycler program used consisted of an initial incubation at 37°C for 30 min and a second step at 95°C for 5 min, after which the annealed oligos were diluted in nuclease-free water (1:200). The vector was digested with BsaI-HF (NEB) and the free overhangs were dephosphorylated with Calf intestine phosphatase (CIP, NEB), to prevent re-ligation of the vector. Phenol-chloroform extraction of the digested plasmid was employed, alternatively to band cutting and DNA extraction from agarose gel. Overnight ligations were prepared for each sgRNA (annealed DNA oligo) at 16°C, using the T4 DNA ligase (NEB). Then, ligation reactions were transformed in Stable Competent *E. coli* (NEB) and grown in Carbenicillin plates to form individual colonies. Finally, 2 to 3 colonies were picked from each plate and the correct insertion of the sgRNAs was verified by digesting with BsaI-HF enzyme. A final confirmation was made by Sanger sequencing using a primer annealing to the U6 promoter, U6F (Table 2.4).

Cloning into the single vector was done sequentially in the same way, although one of the sgRNA cassettes contains a BbsI-cloning site, instead of BsaI site. So, sgRNA1 was inserted using the protocol described above, and the second one was inserted afterwards with the same strategy but using BbsI enzyme (NEB) instead of BsaI-HF. As the U6 primer has two binding sites in this single-vector, Sanger sequencing of cloned oligos was done with SV4F for the upstream sgRNAs, while the downstream ones were sequenced with primers matching the upstream sgRNAs: T4BF, T4IF and Ctrl1F (Table 2.4).

2.2.1.3 Cell culture

HEK293T cells (Human embryonic kidney cell line) were obtained from ATCC and used for screenings of sgRNA activity and AAV productions. HEK293T cells were maintained in Dulbecco's Modified Eagle's Medium (DMEM with 4.5 g/l Glucose, L-glutamine, without sodium pyruvate, Sigma), supplemented with 10% FBS and 100 U/ml of penicillin and 100 µg/ml streptomycin (Lonza).

For primary cortical neuron culture, E16 embryos were dissected from Wild type (WT) pregnant females after crossing with transgenic C9-500 BAC Tg males. DNA was isolated from embryos and genotyped for sex and the presence of the transgene. Further details about this colony and the genotyping are provided bellow (section 2.2.2.1 Breeding and genotyping of C9-500 BAC mice). Transgenic or non-transgenic female embryos were then used for further processing. Briefly, the cortices were dissected and digested in 0.25% trypsin in HBSS without calcium or magnesium (GIBCO) at 37°C for 15 minutes and dissociated manually in triturating medium by using three fire-burnt Pasteur pipettes with successively smaller openings. Dissociated cortical neurons were then plated on poly-D-lysine (Sigma) coated plates at a density of 4'500'000 cells/ plate and maintained in Neurobasal medium (Life Technologies) supplemented with 2% B27 (Life Technologies), 0.5 mM GlutaMax (Life Technologies) and 100 U/ml of penicillin and 100 µg/ml streptomycin (Lonza). Cortical neuron dissection and plating for this project was primarily done by Dr Claudia Bauer with assistance from Dr Christopher Webster or myself.

2.2.1.4 Transfections in HEK293T cells

HEK293T cells of low passage number (18-30 passages) were maintained as described in the previous section (2.2.1.3 cell culture) and were seeded onto 24-well plates or 6-well plates at a density of 75,000 or 300,000 cells per well, respectively. Approximately 24 h after plating, cells in 24-well plates were co-transfected with 250 ng of each plasmid, using Pei as a transfection reagent in a 3:1 ratio (Pei:DNA). Transfections of the single-vectors were done with the same protocol but using 500 ng of the plasmid DNA per well. In 6-well plates, a total of 2 µg of DNA was used per well, also maintaining a 3:1 ratio of Pei:DNA. The cells were harvested 48 hours after transfection and processed for either genomic DNA extraction with Quick-DNA™ Miniprep Kit (Zymo Research) or protein extraction for Western-blot.

2.2.1.5 Amplification of exon 1a, intron 1 and exon 1b of C9orf72

The Primer-BLAST from NCBI (Ye et al., 2012) was used to design the highly specific C9_OUT_7 and C9_OUT_2 primer pairs, spanning the exon 1a, first intron and exon 1b of the human C9orf72 gene (Table 2.4). The first attempts to amplify this genomic region were made through conventional PCR reactions using the FirePol mastermix (Solis BioDyne), 200 nM of each primer and 1ng/µl of gDNA. The thermocycler program consisted in the following: 95°C for 2 min, followed by 30 cycles of 95°C for 20s, 58°C for 20s and 72°C for 30s. A final extension was done at 72°C for 3 min and the PCR product was visualized on a 2% agarose gel. The failure to amplify this region was due to the GC-rich nature of the sequence, which promotes the formation of DNA secondary structures that hamper the progress of the polymerase. To overcome this issue, we added 10% DMSO to the PCR reaction to promote the denaturation of GC-linked secondary structures. After several troubleshooting steps, we came up with the optimal conditions for the amplification of this region of interest (Table 2.6):

Table 2.6. PCR reaction to amplify C9orf72 region of interest. List of PCR reagents and their working concentrations and volumes to amplify exon 1a, intron 1 and exon 1b of C9orf72.

Reagents	Working concentration	Final volume used
ddH2O		to final 50 µl volume
C9_OUT_7/2_Fwd	0.2 µM	1 µl of 10 µM primer stock
C9_OUT_7/2_Rev	0.2 µM	1 ul of 10 µM primer stock
FirePOL 5x RTL 7.5	1x	10 µl of 5x stock
DMSO	10 %	5 µl
DNA	Final quantity: 50-150 ng	To final quantity 50-150 ng

The thermocycler program is the following:

Initial denaturation

95°C 3 minutes

Cycle 35x

95°C 30 seconds

57°C 45 seconds

72°C 40 seconds

Final elongation

72°C 10 minutes

Hold

10°C

The PCR product was then visualized on a 2% agarose gel.

2.2.1.6 PCR amplification and sequencing of G4C2 flanking regions

The intronic sequences flanking the G4C2 region of the C9orf72 gene were PCR amplified in gDNA from 3 different C9-500 BAC mice (Liu et al., 2016). C9_INT_1 and C9_EXT_3 primer pairs (Table 2.4) were designed with Primer-BLAST (Ye et al., 2012) for amplification of the upstream and downstream flanking sequence, respectively. Conventional PCR was used with FirePol mastermix, 200 nM of each primer and 150 ng of template DNA (or 1 µl of QuickExtract gDNA from mouse ear clips). The program used for C9_INT_1 pair was the following:

Initial denaturation

95°C	3 minutes

Cycle 35x

95°C	30 seconds
52°C	45 seconds
72°C	40 seconds

Final elongation

72°C	5 minutes

Hold

10°C

The program used for C9_EXT_3 primer pair was identical, except for the annealing temperature being 56°C instead of 52°C.

Following amplification, PCR products were run on an agarose gel, after which the bands matching the correct size were cut and the DNA was extracted from agarose using the Wizard® SV Gel and PCR cleanup kit (Promega). The purified DNA was then inserted into a TOPO cloning vector, using the TOPO® TA Cloning® Kit (Invitrogen), according to the manufacturer's instructions. Transformed bacteria were grown in Kanamycin plates overnight, after which 2 colonies were picked for each condition to screen for the presence of the insert by digestion of the plasmid DNA with EcoRI (NEB). Positive clones were then sent for Sanger sequencing using the M13_Fwd primer (Table 2.4).

2.2.1.7 Cloning of the pAAV-CRISPR-Cas9 single-vector

In order to clone a single AAV vector expressing SaCas9 with 2 sgRNA cassettes, we designed 2 gBlocks (Integrated DNA technologies) with BsmBI restriction sites in the extremities, compatible with each other and the sub-cloning vector (Table 2.3). The left gBlock contained a U6 promoter, a dual BsaI site for sgRNA1 cloning and a sgRNA scaffold. The right gBlock contained a U6 promoter, a dual BbsI site for sgRNA2 cloning and a sgRNA scaffold sequence. The two blocks were assembled together in the LentiCRISPR V2 plasmid (a gift from Feng Zhang (Sanjana et al., 2014)) by using the golden gate cloning strategy. Then, a conventional cloning step with restriction enzymes NotI and KpnI was used to cut the entire fragment (U6:BsaI-site:sgRNA scaffold:U6:BbsI-site:sgRNA scaffold) and insert it in the pX601 plasmid, replacing the original sgRNA cassette and creating the pX601::dual sgRNA cassette. The correct assembly of the inserts was verified by Sanger sequencing, using 4 sequencing primers: SV3F, SV3R, SV4F and SV5F (Table 2.4) (Figure 2-1).

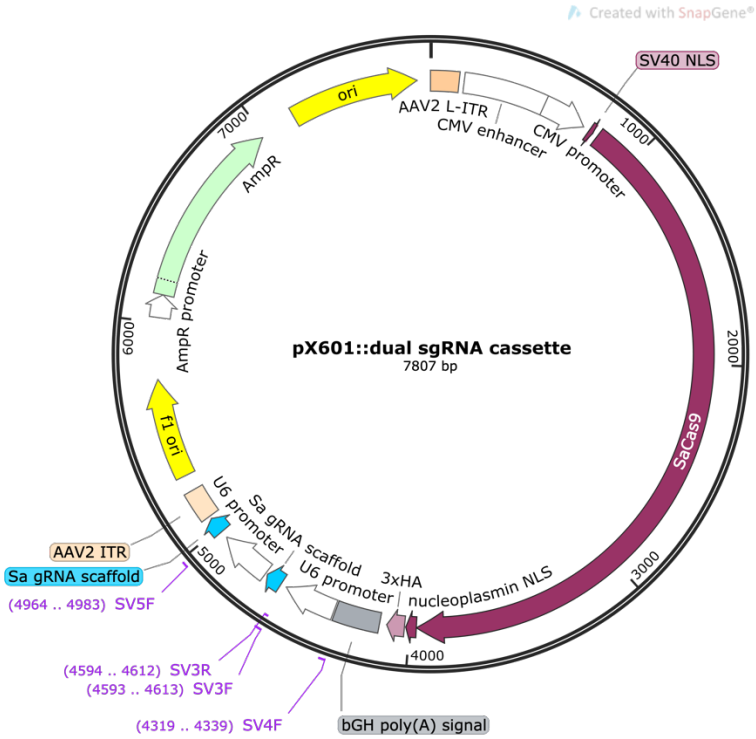


Figure 2-1. Plasmid map of pX601::dual sgRNA cassette. The plasmid map of the cloned AAV construct pX601::dual sgRNA cassette, containing a CMV-driven SaCas9 transgene with a bGH poly(A) sequence followed by a U6:sgRNA cloning site I:sgRNA scaffold:U6:sgRNA cloning site II:sgRNA scaffold double cassette, all flanked by AAV2 inverted terminal repeats (ITRs). The primers used for sequencing verification after cloning are shown in purple. This map was exported from SnapGene viewer, SnapGene® software (from GSL Biotech; available at snapgene.com).

The EFS1a promoter was amplified from the LentiCRISPR V2 by conventional PCR, using Phusion DNA polymerase (NEB) and 200 nM of the EFS1a_Fwd and EFS1a_Rev primers (Table 2.4), containing XbaI and NcoI restriction sites, respectively. The thermocycler program was the following:

Initial denaturation

98°C 3 minutes

Cycle 35x

98°C 10 seconds

67°C 30 seconds

72°C 10 seconds

Final elongation

72°C 10 minutes

Hold

10°C

The PCR product was purified by running on an agarose gel followed by DNA extraction from the cut band, with the GenElute™ Plasmid Miniprep Kit (Sigma). The DNA amplicon

containing the promoter sequence was then digested with XbaI and NcoI (NEB) and loaded again on an agarose gel, followed by band cutting and DNA extraction. Finally, the pX601::dual sgRNA cassette was also digested and purified with the same enzymes, removing the CMV promoter, and was incubated with the EFS1a in a 20 min ligation reaction at room temperature. The ligations were then transformed in stabl3 bacteria (TransGen Biotech) and grown in Carbenicillin plates to form individual colonies. The correct insertion of the EFS1a promoter was verified by Sanger sequencing using the Prm_Rev primer (Table 2.4), forming the pX601::EFS1a::dual sgRNA cassette (Figure 2-2).

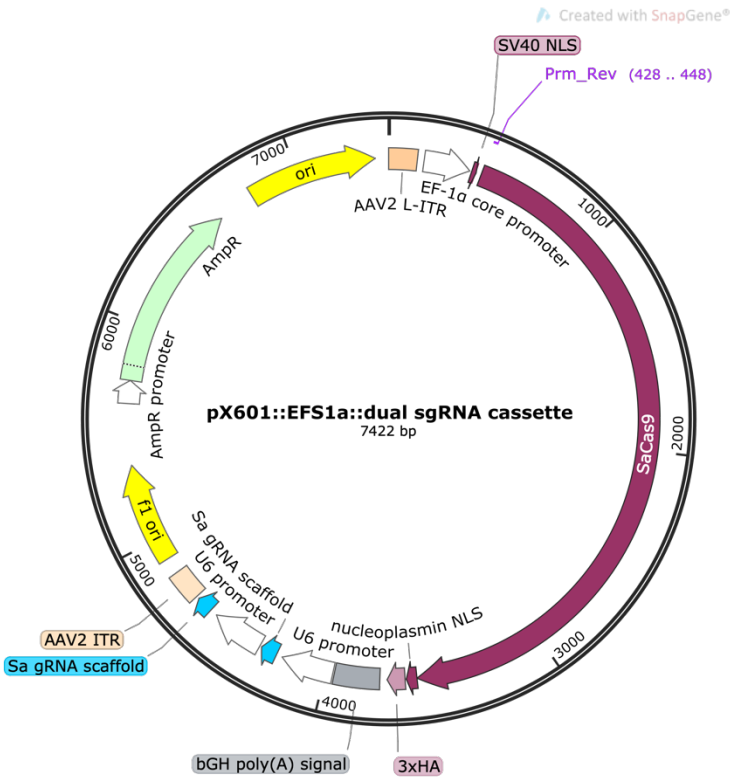


Figure 2-2. Plasmid map of pX601::EFS1a::dual sgRNA cassette. The plasmid map of the cloned AAV construct pX601::EFS1a::dual sgRNA cassette. The main features include an EFS1a-driven SaCas9 transgene with a bGH poly(A) sequence followed by a U6:sgRNA cloning site I:sgRNA scaffold:U6:sgRNA cloning site II:sgRNA scaffold double cassette, all flanked by AAV2 inverted terminal repeats (ITRs). The Prm_Rev primer used for sequencing verification after cloning is shown in purple. This map was exported from SnapGene viewer, SnapGene® software (from GSL Biotech; available at snapgene.com).

The synthetic 49 bp poly(A) signal (Gray et al., 2011; Levitt et al., 1989) was designed as 2 self-complementary DNA oligonucleotides (Table 2.4), whose annealing created sticky ends compatible with EcoRI and KpnI. These 2 restriction sites also flank the bGH poly(A) sequence in the pX601::EFS1a::dual sgRNA cassette plasmid, for which reason a standard "cut and paste" cloning were sufficient to replace it with the sPoly(a). The correct insertion of the sPoly(A) was verified by Sanger sequencing using the sPAF primer (Table 2.4), forming the final configuration of the pAAV-CRISPR-Cas9 single-vector (Figure 2-3).

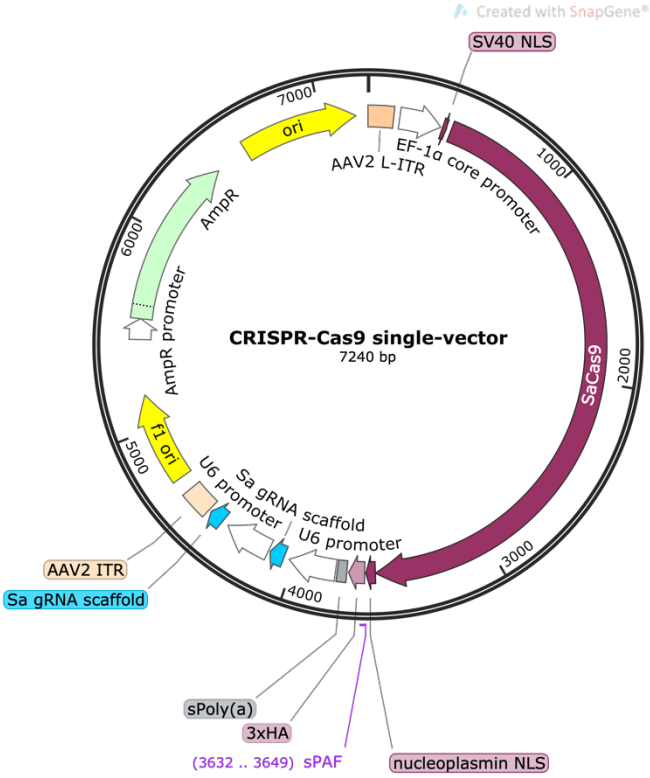


Figure 2-3. Plasmid map of final pAAV-CRISPR-Cas9 single-vector. The plasmid map of the final cloned configuration of the pAAV-CRISPR-Cas9 single-vector. The main features include an EFS1a-driven SaCas9 transgene with an sPoly(A) sequence followed by a U6:sgRNA cloning site I:sgRNA scaffold:U6:sgRNA cloning site II:sgRNA scaffold double cassette, all flanked by AAV2 inverted terminal repeats (ITRs). The sPAF primer used for sequencing verification after cloning is

shown in purple. This map was exported from SnapGene viewer, SnapGene® software (from GSL Biotech; available at snapgene.com).

2.2.1.8 Sequencing of gene-edited bands

Following amplification, with C9_OUT_7 or C9_OUT_2 primers, of the exon 1a, first intron and exon 1b from Cas9-edited HEK cells, the PCR products were run on an agarose gel, after which the top/unedited (~650 bp) and lower/edited (~400 bp for C901 editing and ~540 bp for C904 editing) bands were cut and the DNA was extracted using the GenElute™ Plasmid Miniprep Kit (Sigma). In edited C9-500 BAC mouse cells, DNA previously edited by C901 viral vector was extracted from a single band, at the expected sizes (~400 bp). In C9-500 BAC cells edited by C904, DNA was amplified by PCR using the C904_INT_14 primer pair using the following program:

Initial denaturation

95°C	3 minutes

Cycle 35x

95°C	30 seconds
53°C	45 seconds
72°C	40 seconds

Final elongation

72°C	5 minutes

Hold

10°C

This PCR originated a band ~90 bp smaller than the unedited DNA (~365 bp band) in cells edited with C904 virus. The DNA from the edited band was then purified and inserted into a TOPO cloning vector, using the TOPO® TA Cloning® Kit (Invitrogen), according to the manufacturer's instructions. Transformed bacteria were grown in Kanamycin plates overnight, after which 2 colonies were picked for each condition to screen for the presence of the insert by digestion of the plasmid DNA with EcoRI (NEB). Positive clones were then sent for Sanger sequencing using the M13_Fwd primer (Table 2.4).

2.2.1.9 PCR-based methods for editing efficiency

We designed a qPCR assay based on previously published work (Goyenvalle et al., 2012, 2015; Tabebordbar et al., 2016) to quantify precision editing efficiency of our therapeutic strategies, which we defined as editing events that resulted in the complete excision of the targeted sequences and full re-ligation of the C9orf72 gene.

For the C901 editing strategy (introduced in chapter 1.5), we designed qPCR primers to bind directly to the predicted re-ligation junction site of C901 (GE_Cut primer pair, Table 2.4), which will only amplify precision-edited C9orf72 molecules. A different pair of primers were designed to bind ~80 bp upstream of this ligation site (GE_Ctrl primer pair, Table 2.4), which amplifies the total number of C9orf72 molecules in a pool of DNA, edited or non-edited. The delta-C_t values for each sample were then used to calculate a percentage of precision-edited C9orf72, relative to the total C9orf72 molecules amplified. This assay assumes that SaCas9 cuts the target DNA always in the same exact position, and that the C9orf72 gene re-ligates always in the same way, which is not always the case. For this reason, this assay always gives an underestimation of true C901 precision editing.

As for the C904 strategy (introduced in chapter 1.5), designing high quality qPCR primers against the re-ligation site proved unfeasible due to the complex nature of that DNA sequence, as indicated by Primer-BLAST (Ye et al., 2012). Therefore, we instead did endpoint PCR with C904_INT_14 primers as described above (see 2.2.1.8 Sequencing of gene-edited bands) and quantified the ratio of edited:total band intensity using Fiji (Schindelin et al., 2012). We then

calculated the relative abundance of edited molecules to the total amount of C9orf72 (Figure 2-4), a method used by other authors as well (Yoshioka et al., 2015), and plotted the results as percentage of C904-mediated precision editing.

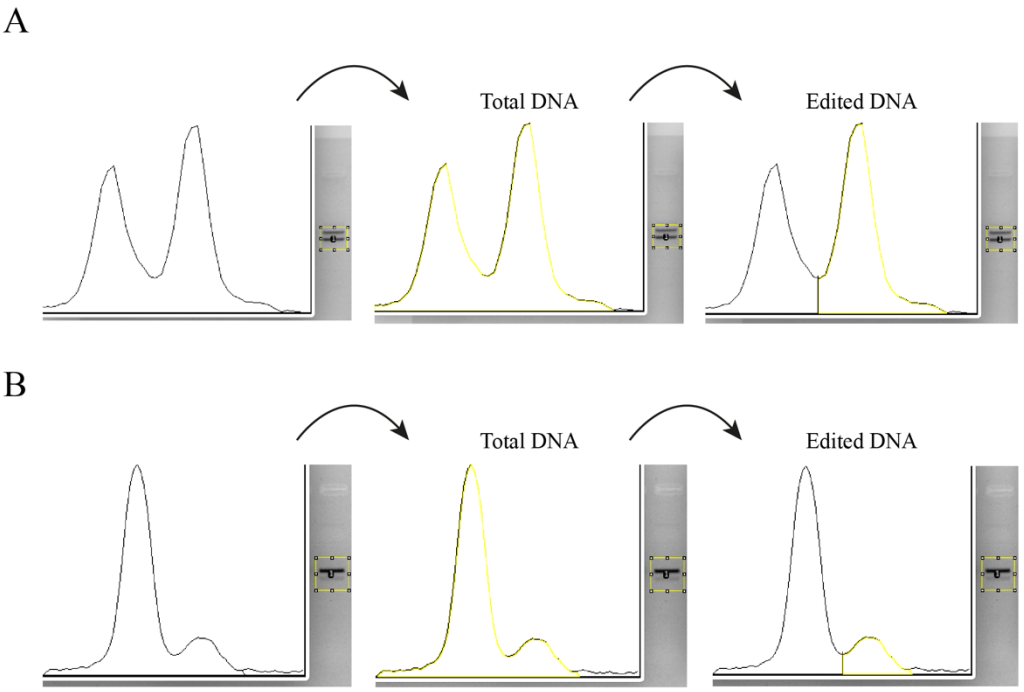


Figure 2-4. Quantification method of C904-mediated editing. This figure shows two example sequences of the method used to quantify C904-mediated precision editing. The two examples show the application of the method in (A) a case with robust pC904 editing in HEK293T cells and (B) lower editing efficiency by C904 in AAV transduced C9-500 BAC cortical neurons. Endpoint PCR of C904-edited DNA with C904_INT_14 primers yields two amplification products (left-hand images), as discussed above (see 2.2.1.8 Sequencing of gene-edited bands). We first quantified the total intensity of both bands (middle images), corresponding to the total amount of PCR product, followed by quantification of the edited band alone (right-hand images). This was then used to calculate relative abundance of edited C9orf72 and the percentage of precision editing mediated by the C904 strategy.

2.2.1.10 Western blotting

Unless stated otherwise, total cellular protein was extracted using the following protocol: cells were washed with ice-cold PBS and lysed with a nuclear and cytoplasmic RIPA (radioimmunoprecipitation) lysis buffer [5% Tris-HCl (pH 7.4), 1% NP-40, 0.5% Sodium deoxycholate, 0.01% SDS, 150 mM NaCl, 0.2 mM EDTA] supplemented with 1% Protease Inhibitor Cocktail (Sigma). Protein concentration was determined by BCA (Bicinchoninic acid) protein assay (Thermo Scientific) according to the manufacturer's instructions using PHERAstar FS spectrophotometer plate reader (BMG Labtech) followed by denaturation in SDS protein sample buffer (10% Glycerol, 60 mM Tris/HCl pH 6.8, 2% SDS, 0.01% bromophenol blue, 1.2% beta-mercaptoethanol) by heating at 95°C for 5 minutes. The same amount of protein from each sample was then loaded into a 10% SDS-polyacrylamide gel and electrophoresed in Tris/Glycine/SDS running buffer at 60V for 15 minutes then 1-1.5 hours at 100 V. For immunodetection, proteins were transferred onto a Polyvinylidene difluoride (PVDF) membrane in cold transfer (Tris/Glycine/Methanol) buffer at 250mA for 1.5 hours. Blocking of the membrane was done in 5% milk in TBST (0.137 M NaCl, 25.92 mM Tris base, 0.1% Tween20) (Sigma) and antibodies mouse anti-HA (1:10'000, Cell signalling), rabbit anti-C9orf72 (1:1000, Proteintech, 25757-1-AP), mouse anti-GAPDH (1:3000, Merck Millipore), chicken anti-GFP (1:5000, Abcam) or rabbit anti-VP1, VP2, VP3 (1:400, ARP) were incubated in blocking buffer overnight at 4°C. The proteins were visualized using the ECL Plus chemiluminescence detection kit (GE Healthcare) and images were captured using the G: BOX gel imaging system (Syngene). The band intensities were quantified using Fiji software (Schindelin et al., 2012). The Western blots were optimised and done by either myself or Ms Sneha Simon. The HEK C9orf72 KO cell line was created at SITraN by Dr Christopher Webster.

2.2.1.11 AAV production

Adeno-associated viruses serotype 9 (AAV9) were packaged by HEK293T cells following Pei transfection of 3 different plasmids, with the same protocol used routinely in our team (Herranz-Martin et al., 2017). Briefly, HEK293T were plated at 80-90% confluency into either 30 or 60 cell culture dishes (150 mm) and left for 24 h, after which they were incubated in serum-free medium and co-transfected with the following plasmids, per dish: 13 µg of the cloned pAAV constructs, containing the ITR-flanked transgenes; 13 µg of the AAV Rep-Cap proteins (pAAV2/9); and 26 µg of the pHelper plasmid, which provides helper genes isolated from adenovirus. Viral particles were harvested at day 5 post-transfection by collecting the medium from all dishes, incubating with 12.5 units/ml of benzonase nuclease (Sigma) for 2-3 hours and filtering the solution in a 0.22 µm filter. The virus was then concentrated using Amicon spin filter units (Millipore) to a final volume of 28 ml and purified in a discontinuous iodixanol (Sigma, D1556) gradient (4 ml of 54%, 9 ml of 40%, 9 ml of 25%, 5 ml of 15%), with centrifugation at 69,000 rpm for 1.5 h at 18°C. Phenol Red was added to the 54% and 25% iodixanol layers to identify the cleared virus-enriched layer. After centrifugation, viral particles were removed into 0.5 ml fractions using a 19-Gauge needle, after which 6 µl of each fraction were mixed at an equal ratio with 2X Laemmli sample buffer, heated to 95°C for 5 min, separated on a poly-acrylamide gel, and stained with Sypro-Ruby according to the manufacturer's instructions (Life Technologies). Fractions that showed a pure virus composed exclusively of the VP1, VP2 and VP3 bands were pooled together, and washed several times in PBS supplemented with 35 mM NaCl in successive centrifugations with an Amicon filter, before collecting in a final volume of 300-500 µl and frozen at -80°C for long-term storage. All AAV9 viral productions were made in-house by me, except for one batch of C904 AAV9 virus, outsourced to the Children's Hospital of Philadelphia (CHOP) research vector core (Table 2.2).

For AAV PHP.eB production, we followed a detailed protocol published elsewhere (Challis et al., 2019), which shares similarities to the protocol we used for AAV9 but also some differences, which we will highlight next. Briefly, HEK293T were plated at 80-90% confluency in up to 10 cell culture dishes (150 mm) per viral production. They were transfected 24 h later with the following plasmids, per dish: 5.7 µg of pAAV construct, 22.8 µg of PHP.eB capsid, and 11.4 µg of pHelper (Table 2.1), following a 1:4:2 ratio. The medium from the cells

was harvested at 3 days post-transfection, and then again at 5 days, together with the cells. Viral particles in the cells were extracted by lysing the cells in a 500 mM NaCl, 40 mM Tris, 10 mM $MgCl_2$ buffer and digesting released nucleic acids with a salt-activated nuclease (SAN, Arcticzymes) at 37°C for 30 minutes. Viruses in the collected medium were precipitated with 40% polyethylene glycol (PEG, Sigma, 89510-1KG-F), after which the PEG pellet was resuspended, digested with SAN, and combined with the cell extracts. Debris from this precipitate was then clarified by centrifugation and the supernatant was loaded in an iodixanol gradient (Optiprep, Sigma; D1556) to isolate viral particles from contaminating proteins. Finally, Amicon filters (EMD, UFC910024) were used to wash and concentrate the viral fraction, which was then stored at 4°C in a PBS solution. All PHP.eB viral productions were done in-house by me (Table 2.2).

2.2.1.12 AAV9 viral capsid integrity

To check for AAV9 capsid integrity, 5 µl of each virus were mixed at an equal ratio with Laemmli sample buffer, heated to 95°C for 5 min and separated on a poly-acrylamide gel. The denatured and separated viral capsid proteins were then processed as a normal Western blot, described above, and incubated with an antibody targeting the AAV9 viral capsid proteins: rabbit anti-VP1, VP2, VP3 (1:400, ARP). At this stage, successfully purified viral productions with functional capsids will show three distinct bands at 87 KDa (VP1), 73 KDa (VP2) and 62KDa (VP3). This step was not done for AAV PHP.eB, as it was not required in the protocol we followed (Challis et al., 2019).

2.2.1.13 AAV titration

Viral titers from both AAV9 and AAV PHP.eB were determined with the Quantifast SyBR Green PCR Kit (Qiagen, 204054) on a Bio-Rad CFX96 thermocycler, following the

manufacturer's instructions, using primers hybridizing in the Cas9 cDNA (Cas9_Fwd and Cas9_Rev, Table 2.4). For each virus, a standard curve was generated by serial dilutions of the corresponding transgene plasmid, previously linearized. A linear regression fit of the standard curve was then used to calculate the number of copies in three dilutions of the purified AAVs (10^{-2}, 10^{-3}, 10^{-4}) and an average of the 3 values was used as the final titer.

2.2.1.14 Immunocytochemistry

Cortical neurons cultured on glass coverslips were fixed in 4% PFA for 10 minutes, washed with PBS and permeabilized with 0.5% Triton X-100 in PBS. Cells were then blocked with 3% BSA in PBS for 30 minutes at room temperature and incubated with primary antibodies diluted in blocking buffer for 2 hours at room temperature. The primary antibodies utilised in this study included anti-MAP2 (1:500, Sigma-Aldrich) and anti-HA (1:1000, Cell signalling). Finally, the cells were counter-stained with secondary Alexa-Fluor® conjugated antibodies (1:1000, Invitrogen) diluted in blocking buffer and incubated for 1 hour at room temperature. Hoechst stain was used to visualize nuclei. Images were captured using a Leica SP5 confocal microscope and quantifications were done manually and blinded to the experimental conditions, using the Fiji software (Schindelin et al., 2012). Immunocytochemistry was done by Dr Evangelia Karyka, who also acquired images. Transduction and quantification of transduced cells were done by me.

2.2.1.15 Repeat-primed PCR

Repeat-primed PCR was done using a protocol adapted from previous work (DeJesus-Hernandez et al., 2011; Hantash et al., 2010). Briefly, DNA from C9-500 BAC cells was used as template for a PCR reaction with 5M betaine, OneTaq PCR Master Mix GC Buffer (NEB),

One Taq High GC Enhancer (NEB) and a mix of three primers: MRX-F, MRX-M13R, MRX-R1 (Table 2.4). The cycling program was the following:

94°C 3 minutes

Touchdown cycle x32, -0.5°C per cycle

94°C 1 minute

71°C 45 seconds

68°C 3 minutes

Cycle 30x

94°C 1 minute

53°C 45 seconds

68°C 2 minutes (+10s per cycle)

Final elongation

68°C 5 minutes

PCR products were then separated by capillary electrophoresis and results were analysed for the presence of a sawtooth tail pattern, characteristic of DNA containing a G4C2 repeat expansion. Cells and DNA were processed by me and sent to the Sheffield Diagnostic Genetics Service in the Sheffield Children's Hospital, where an operator blinded to the experiment performed the PCR and analysis.

2.2.1.16 Fluorescence in situ Hybridization (FISH)

FISH was performed following the same protocol described previously by our research group, with minor changes (Cooper-Knock et al., 2014a; Walker et al., 2017). Mouse cortical neurons were fixed with 4% PFA for 20 minutes at room temperature. Firstly, cells were incubated with hybridisation buffer (50% formamide, 2x saline sodium citrate (SSC), 100 mg/ml dextran sulphate, 50 mM sodium phosphate pH 7.0) for 1 hour at 66°C. They were then incubated with hybridization buffer containing a 5′ TYE-563-labelled LNA (16-mer fluorescent)-incorporated DNA probe designed against the sense RNA hexanucleotide repeat (1:400, Exiqon, batch number 607323) at 66°C overnight. After hybridization, wells were washed once in 2×SSC/0.1% Tween-20 at room temperature (5 minutes) and three times in 0.1× SSC at 66°C, 20 minutes each wash. Finally, cells were counterstained with Hoechst, washed three times in PBS and were visualized in a high throughput confocal microscope (Opera Phenix, Perkin Elmer). All the solutions used in this protocol were treated with Diethyl pyrocarbonate (DEPC *BioChemica*, Panreac Applichem) to eliminate contaminating RNAse enzymes. Quantifications of RNA Foci(+) neurons were done automatically using the Harmony software (Perkin Elmer), which detects nuclear inclusions in an unbiased manner.

2.2.1.17 Dot blotting

Total protein was extracted from cortical neurons, transduced or non-transduced with AAVs, into ice-cold lysis buffer (50 mM Hepes pH7.5, 150 mM NaCl, 10% glycerol, 0.5% Triton X-100, 1 mM EDTA, 1 mM DTT, protease inhibitor cocktail (Sigma)). Protein extracts were then left to lyse on ice for 10 min followed by centrifugation at 17,000 g at 4°C for four minutes. Protein concentration was measured using Bradford Reagent (BioRAD) and 80-100 µg of each extract were loaded onto a nitrocellulose membrane using a microfiltration apparatus (Biorad). The membranes were then sliced into strips and analysed by immunoblotting, using antibodies rabbit anti-poly-Gly-Pro (1:1000, Proteintech) and mouse anti-tubulin (1:10’000, Sigma).

2.2.2 *In vivo* experimental methods

All *in vivo* experiments were approved by the University of Sheffield Ethical Review Sub-Committee, the UK Animal Procedures Committee (London, UK) and performed according to the Animal Scientific Procedures Act 1986, under the Project License 40/3739. C9-500 mice were purchased from The Jackson Laboratory (stock 029099) and were maintained in a controlled facility in a 12 h dark/12 h light photocycle (on at 7 am/ off at 7 pm) with *ad libitum* access to food and water. All animal behavioural experiments in this project were done while blinded to their treatments.

2.2.2.1 Breeding and genotyping of C9-500 BAC mice

C9-500 is a Bacterial Artificial Chromosome (BAC)-transgenic strain on a FVB/NJ background, which carries the entire human C9orf72 sequence with approximately 500 hexanucleotide repeats, plus a significant portion of sequence upstream (~50 Kb) and downstream (~20 Kb) of the gene (Liu et al., 2016). This expanded allele was isolated from a patient-derived lymphoblastoid cell line (LCL). For expanding and maintaining the colony, hemizygous males were crossed with WT FVB/NJ inbred females and litters were ear-clipped at approximately 2 weeks of age for genotyping by PCR, using a protocol adapted from The Jackson Laboratory. Briefly, genomic DNA was extracted from the ear clips using QuickExtract (DNA Extraction Solution 1.0, Epicentre Biotechnologies) according to the manufacturer's instructions. Then, 0.5 µl of the isolated DNA was used as a template for two PCR reactions (Tables 2.7 and 2.8):

Table 2.7. Internal control reaction. List of reagents and their working concentrations and volumes to amplify the C9-500 BAC genotyping internal control product.

Reagents	Working concentration	Final volume used
ddH2O		6.5 µl
21238	0.5 µM	0.5 µl of 10 µM primer stock
21239	0.5 µM	0.5 µl of 10 µM primer stock
FirePOL 5x RTL 7.5	1x	2 µl of 5x stock
DNA	Final quantity: 50-200 ng	0.5 µl

Table 2.8. Transgene reaction. List of reagents and their working concentrations and volumes to amplify the transgene for C9-500 BAC genotyping.

Reagents	Working concentration	Final volume used
ddH2O		6.5 µl
35766	0.5 µM	0.5 µl of 10 µM primer stock
35767	0.5 µM	0.5 µl of 10 µM primer stock
FirePOL 5x RTL 7.5	1x	2 µl of 5x stock
DNA	Final quantity: 50-200 ng	0.5 µl

Both reactions were ran simultaneously under the following conditions:

Initial denaturation

94°C 2 minutes

Touchdown cycle x10, -0.5°C per cycle

94°C 20 seconds

65°C 15 seconds

68°C 10 seconds

Cycle 28x

94°C 15 seconds

60°C 15 seconds

72°C 10 seconds

Final elongation

72°C 2 minutes

Hold

10°C

At the end of the reaction, the PCR products were loaded onto a 2% agarose gel with 0.5 µg/µl ethidium bromide and subjected to electrophoresis for 25 minutes at 120 V in a standard TAE buffer (40 mM Tris, 20 mM Acetate and 1 mM EDTA). Gels were visualised on a GENi imaging system (Syngene). All animals were expected to have an internal control band at 415 bp whereas only transgenics show a band at 213 bp.

For cortical neuron experiments, embryos were also genotyped for sex, using the following PCR reaction (Table 2.9):

Table 2.9. Sexing by PCR. List of reagents and their working concentrations and volumes for the sexing of mice by PCR.

Reagents	Working concentration	Final volume used
ddH2O		18.5 µl
sex 1 sry_Fw	0.12 µM	1 µl of 3 µM working solution
sex 1 sry_Rev	0.12 µM	1 µl of 3 µM working solution
FirePOL 5x RTL 7.5	1x	5 µl of 5x stock
DNA	Final quantity: 50-200 ng	0.5 µl

Sexing PCR reactions were run with the following program:

Initial denaturation

95°C 10 minutes

Cycle 30x

95°C 30 seconds

65°C 30 seconds

72°C 30 seconds

Final elongation

72°C 10 minutes

Hold

10°C

After the cycling program is over, the reactions were resolved on a 2% agarose gel. A single band is expected at approximately 200bp in males while no band is expected in females. Genotyping of C9-500 BAC mice was optimized by myself and Mr Ian Coldicott and then done routinely by Mr Ian Coldicott.

To maintain reproducibility of the *in vivo* experiments, we investigated whether all male breeders were capable of transmitting a fully functional G4C2 expansion to their offspring. We did this by analysing RNA Foci expression in embryonic cortical neurons cultured from their progeny. We noticed that some male breeders' offspring did not show, or showed a very reduced number of, RNA Foci accumulation. We hypothesised that this could be due to contractions in the number of G4C2 repeats, or hypermethylation of the promoter in key areas. Therefore, all animals used in our *in vivo* studies originated from the same breeders that were previously confirmed to produce progeny with significant RNA Foci expression.

2.2.2.2 Viral delivery to mice

For tail vein injections, animals aged 3-4 weeks old were placed in a warmer environment (31°C) for up to 15 minutes and then firmly held with the aid of a restraining device. A heat lamp was used to further dilate the lateral veins in the tail, after which mice received a single intravenous dose of $1x10^{12}$ vg of AAV9 per mouse, in a final volume of 100 µL. For animals injected with AAV PHP.eB, mice were injected at approximately 6 weeks old and received a total of $2x10^{11}$ vg per mouse in a final volume of 80 µL.

2.2.2.3 Dissection of mouse tissue for DNA and protein collection

Mice were given an overdose of pentobarbital and were transcardially perfused with 20 ml of room temperature PBS followed by dissection of the entire body. To obtain a more complete picture of how widespread the viral biodistribution and *in vivo* editing was throughout the CNS, the brain was divided into frontal lobe, parietal-temporal lobes and cerebellum (Dorr et al., 2007), while the spinal cord was divided into cervical, thoracic and lumbar cord. Peripheral tissues were also collected including heart, liver and hindlimb muscle. All the tissue was immediately snap-frozen in liquid nitrogen upon collection, after which it was stored at -80°C until further processed.

2.2.2.4 Viral biodistribution

Genomic DNA was extracted from frozen tissue using Quick-DNA™ Miniprep Kit (Zymo Research) for CNS tissue and the GenElute™ Mammalian Genomic DNA Miniprep Kit (Sigma) for heart, liver and muscle tissue. To calculate viral genome load in different organs

of the mouse, we did the exact same qPCR assay used to titer AAV productions (see 2.2.1.13 AAV titration). The results were plotted as viral genomes per ng of extracted DNA (vg/ng).

2.2.2.5 In vivo gene editing

Genomic DNA was extracted from frozen tissue using Quick-DNA™ Miniprep Kit (Zymo Research) for CNS tissue and the GenElute™ Mammalian Genomic DNA Miniprep Kit (Sigma) for heart, liver and muscle tissue. C901 or dual-vector C901 (dvC901)-mediated gene editing was assessed by doing PCR from extracted DNA with C9_OUT_7/2 primer pairs, optimised previously (see 2.2.1.5 Amplification of exon 1a, intron 1 and exon 1b of C9orf72). This PCR amplifies a single ~400bp band in DNA edited with the C901 strategy.

To study C904 or dual-vector C904 (dvC904)-mediated editing, PCR with C904_INT_14 primer pair was done, as described above (see 2.2.1.8 Sequencing of gene-edited bands). Similarly to the *in vitro* data, this PCR originates a band ~90 bp smaller than the unedited DNA (~365 bp band) in cells edited with the C904 strategy.

2.2.2.6 Poly-(GP) ELISA of mouse brains

A portion of the parietal-temporal lobes of mouse brains was mechanically homogenised and allowed to lyse on ice for 10 minutes in a RIPA lysis buffer [150 mM NaCl, 1.0% NP-40, 0.5% sodium deoxycholate, 50 mM Tris, pH 8.0] supplemented with 1% Protease Inhibitor Cocktail (Sigma) and 2% SDS. Lysates were then centrifuged at 17'000 g for 20 minutes at room temperature and supernatant was collected into tubes. Protein concentration was determined by BCA (Bicinchoninic acid) protein assay (Thermo Scientific) according to the manufacturer's instructions using PHERAstar FS spectrophotometer plate reader (BMG Labtech), after which lysates were all diluted to the same concentration of 1 mg/ml. Samples

were then sent to the lab of Professor Adrian Isaacs at UCL, where they were blindly analysed on an MSD imager following a previously published protocol (Simone et al., 2018).

2.2.2.7 Body weight

Due to the high numbers of animals, body weight was routinely taken once every two weeks after treatment administration, to monitor any adverse reactions to the viruses, and later on to spot onset of symptoms. However, sometimes weight was temporarily monitored on a daily basis, as soon as a mouse showed any signs of distress.

2.2.2.8 Open-field

The open field test was done as previously described by our team (Herranz-Martin et al., 2017), across 3 different timepoints: 10-12 weeks, 18-20 weeks and 26-28 weeks. Animals were gently placed on a semi-transparent plastic box measuring 60 cm×40 cm×25 cm with a grid of 3×5 squares drawn on the bottom with a permanent marker. Activity was measured for 10 minutes in a room with minimal light, during which the number of squares crossed was recorded. Before the beginning of the study, it was previously defined that a mouse crossed a square when all four paws crossed over one of the boundaries.

2.2.2.9 Grip strength

The grip strength of mice was measured at 35-37 weeks, 38-40 weeks and 40-42 weeks with a four limb hang test, adapted from previous studies with mouse models (Aartsma-Rus and van Putten, 2014; Deacon, 2013; Gleitz et al., 2017; Kondziella, 1964). Mice were placed on a wire grid system, which was then inverted to force the animals to sustain their own weight using just their limb tension for up to 2 minutes. For mice that fell off the grid before the maximum time, they were given 2 extra tries and the best score was used for analysis.

2.2.2.10 Electrophysiology-CMAP

In preparation for *in vivo* electrophysiological recordings, mice were first placed under a gaseous anaesthesia (maintained with a gas mixture of oxygen and 1–2% isoflurane through a nose cone at a flow rate of 1 L/min) and body temperature was maintained using an electric heat pad (CWE, USA). Also, the fur of the left hindlimb was shaved thoroughly to optimise signal conductance.

A Dantec Keypoint Focus EMG System (Optima, UK) was used for the setup. Briefly, a grounding electrode was placed in the base of the tail (Ambu Neuroline, UK), and ring recording electrodes were gently coated with Ten20 conductive paste (Pulse Medical Ltd, UK) and placed circumferentially around the distal hindlimb muscles. A pair of subdermal electrodes (Ambu Neuroline, UK) were then used to subcutaneously apply a square wave electrical impulse of 0.5-1ms duration at the sciatic notch, and the compound muscle action potential (CMAP) amplitude was measured. To ensure we recorded the supramaximal response, we positioned the subcutaneous electrodes so that a response was first obtained using a stimulating current of 1-2 mA, after which the current intensity was increased until no further increase in amplitude was seen. The supramaximal amplitude recording of each animal was used to plot the results. CMAP recordings were taken with assistance from Dr James Alix and Mr Ian Coldicott.

2.2.2.11 Dissection and cutting of mouse brains for histology

Mice were given an overdose of pentobarbital and were transcardially perfused with 20 mL of room temperature PBS followed by dissection of the brain and spinal cord. Ice-cold 4% paraformaldehyde (PFA) was used for post-fixation of the brain, cervical and lumbar spinal cord, to be further analysed by histology. Post-fixation was done for 24 hours, after which the tissue was cryoprotected in a PBS solution containing 30% sucrose until the tissue sank. At this point, brains and spinal cords were embedded in optimal cutting temperature (OCT) and frozen at -80°C until ready to be sectioned. Hindlimb muscle was also post-fixed for posterior analysis of Neuromuscular junction (NMJ).

Mouse brains were sliced in a cryostat, with sections either cut for free-floating with 40 μm thickness, or cut directly onto microscope slides at 16 μm thickness.

2.2.3 Statistical analysis

Unless otherwise stated, all data are presented as mean ± standard error of the mean (s.e.m.) of at least 3 experimental replicates. Statistical analysis was performed using GraphPad Prism 8. Statistical differences were analysed using the following tests: One-way Anova; Chi-square test; Mantel-Cox (log-rank) test, Gehan-Breslow-Wilcoxon test; with $p < 0.05$ considered to be statistically significant.

3. Design and validation of gene editing therapeutic vectors

3.1 Introduction

As detailed previously (chapter 1.5), this project focuses on designing a highly targeted gene therapy to treat C9orf72-linked ALS and FTD which, despite vast pre-clinical research (Donnelly et al., 2013; Jiang et al., 2016; Lagier-Tourenne et al., 2013; Martier et al., 2019a, 2019b; Sareen et al., 2013; Simone et al., 2018), does not have any market-approved treatment. In this project, we decided to counter the considerable complexity of C9orf72-linked neurodegeneration by employing one of the most powerful and versatile tools in molecular biology in recent years: CRISPR-Cas9 editing. By the time this project was being planned, three different groups reported the use of CRISPR-gene editing to treat Duchenne muscular dystrophy, using Cas9-mediated dual double-strand cutting to promote the skipping of exon 23 (Long et al., 2016; Nelson et al., 2016; Tabebordbar et al., 2016). Inspired by these results, we wondered whether a simultaneous dual double-strand cut could be used to introduce strategic deletions in the C9orf72 which could modulate its transcription and translation, altering the course of disease. If successful, our first idea was to then use this system to excise the entire repeat expansion from the C9orf72 first intron, whose treatment we later designated C901. However, our main concern with this approach was whether the size of the expansion would complicate its removal and the re-ligation of the gene (from here on referred to as precision editing), instead promoting the localized repair of each double-strand cut without excising the whole fragment. Unsure whether this would constitute a problem, we devised another strategy to test in parallel. The work of two different teams on studying the mechanisms of RAN translation in C9orf72 was crucial to this alternative idea (Green et al., 2017; Tabet et al., 2018). In these publications, the authors reported the presence of a strong Kozak consensus sequence located just 24 bp upstream of the expansion which contains an alternative ‘CUG’ start codon used by the cell to initiate non-canonical RAN translation of the repeats. We hypothesized that introducing a partial deletion in this highly regulatory region had the potential to inhibit RAN translation and DPR production, whose role in C9orf72-mediated pathology has numerous and compelling evidence (Freibaum and Taylor, 2017). We designated this alternative treatment as C904, an approach expected to enable more efficient precision editing by aiming for a shorter deletion, compared to C901.

This chapter describes the first steps and all the considerations we took to translate these ideas from just theoretical concepts into experimental gene therapies ready to be tested pre-clinically.

To achieve this, we broke down this task into smaller, key milestones. First, testing the general concept of the project. In particular, checking whether and how expressing Cas9 and a pair of sgRNAs hybridising in the human C9orf72 could induce programmable deletions in a reproducible manner. Second, considering how to express this system from a clinically relevant single vector, that can be used to produce therapeutic AAV particles in a relatively cost-effective way. Finally, investigating how editing the C9orf72 gene can impact on the levels of endogenous protein, to have an early understanding of potential deleterious effects from these experimental therapies.

3.2 Co-transfection of two AAV CRISPR constructs can edit the human C9orf72 gene: a proof-of-principle

The ultimate aim of this project was to develop a gene therapy that could be delivered from an AAV virus, due to their excellent safety profile and their recent successes in clinical trials which, in some cases, led to market licensing (Carpentier et al., 2012; Gaudet et al., 2013; MacLaren et al., 2014; Mendell et al., 2017; Stroes et al., 2008). Considering this, we decided to use SaCas9 which, after its discovery, was repurposed for AAV-mediated delivery due its smaller size, when comparing with the canonical SpCas9 (Ran et al., 2015). To target the C9orf72 gene with efficient sgRNAs and minimal off-target cutting activity, we took advantage of the freely accessible software Breaking-Cas (Oliveros et al., 2016), which can be used to design the customized target-specific spacer portion of sgRNAs (used interchangeably with the term sgRNA throughout this document, for convenience). Moreover, this software allows selection for customized PAM sequences such as the 'NNGRRT', the optimal PAM for SaCas9 (Ran et al., 2015), used to design every sgRNA in this project. We used ENSEMBL (Zerbino et al., 2018) to retrieve the sequence of the human C9orf72 gene, which we then used as a template. We designed a total of six different sgRNAs on the basis of least off-target effects, with two of them hybridizing upstream of the Kozak sequence (T4B and T4I, Figure 3-1), one binding between the Kozak sequence and the G4C2 repeat region (T4D, Figure 3-1) and three sgRNAs targeting the region downstream of the repeats (Figure 3-1). To screen for the

expected genomic deletions in the C9orf72 gene, we used primer blast (Ye et al., 2012) to design highly specific primers to the 5' region of the gene, encompassing the exon 1a, first intron containing the G4C2 region and the short exon 1b (red arrows, Figure 3-1).

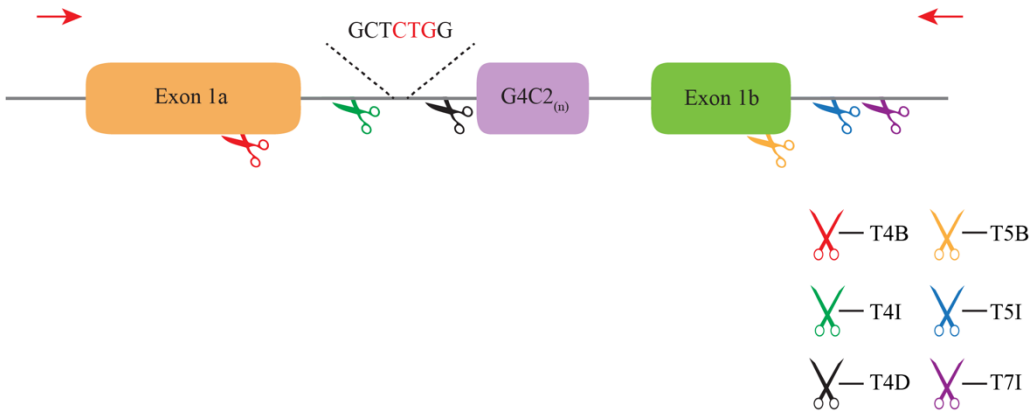


Figure 3-1. sgRNAs targeting the human C9orf72 gene. Schematic outline of the initial portion of the C9orf72 gene, with the hexanucleotide repeat region between the non-coding exons 1a and 1b. The strong Kozak consensus sequence containing the alternative start codon (in red) for RAN translation is shown just upstream of the G4C2 repeat region. The scissors represent the binding sites for the different sgRNAs designed and tested in this project. Expression of the sgRNAs, comprising the customized target-specific spacer and the sgRNA scaffold, will guide the Cas9 nuclease to the targeted regions in the C9orf72. Red arrows represent approximate C9_OUT primer pairs binding sites.

For a quick and preliminary platform to check the validity of our ideas, we decided to use HEK293T cells, which have 3xG4C2 repeats, mainly due to the highly optimised and available protocols to edit their genome with CRISPR (Ran et al., 2013a). Also, their fast-growing rate and easiness to transfect make them an ideal, quick screening platform. We employed a paradigm of co-transfections using an expression construct designed by Ran and colleagues (Ran et al., 2015), pX601, which expresses SaCas9 from a CMV promoter followed by a U6-driven sgRNA cloning site (Figure 3-2). After cloning all sgRNAs in this construct, we were then able to co-transfect HEK293T cells with different combinations of the same vector cloned with the different guides binding upstream and downstream of the repeat region.

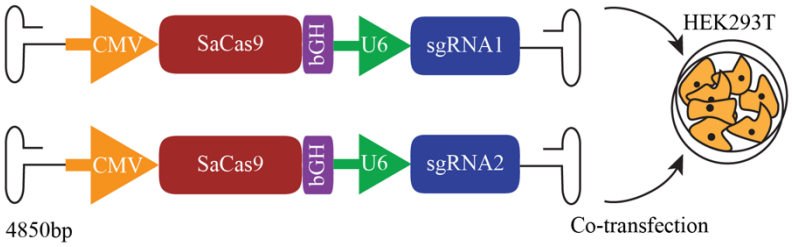


Figure 3-2. Co-transfection of two AAV CRISPR constructs can express different sgRNAs combinations targeting human C9orf72. Schematic representation of the co-transfected AAV pX601 vectors, each expressing *Staphylococcus aureus* Cas9 (SaCas9) driven by the CMV promoter and a U6-driven sgRNA, flanked by ITR sequences. One of the plasmids expresses a sgRNA binding upstream while the second expresses a sgRNA binding downstream of the target sequence. Equal quantities of each plasmid were used for co-transfection.

The cells were incubated for at least 48 hours following co-transfection, after which we performed endpoint PCR on extracted gDNA using C9_OUT primer pairs. This PCR yielded a band of the expected size corresponding to the full, unedited fragment of the C9orf72 gene in every treated and untreated condition (Figure 3-3). Interestingly, however, the same PCR done in gDNA from cells co-transfected with the CRISPR constructs also yielded a lower band (white arrowheads, Figure 3-3), corresponding to the same segment of the C9orf72 without the fragments cut out by our strategy. Since each sgRNA binds to a slightly different sequence, the combinations of co-transfections introduce smaller or bigger deletions, explaining why the edited bands never have the exact same size. In each treated condition, a middle band appears between the unedited and the edited band, a common phenomenon in PCRs from mixed pools of DNAs, corresponding to the heteroduplex formed between those two bands, which migrates at an intermediate speed in an agarose gel.

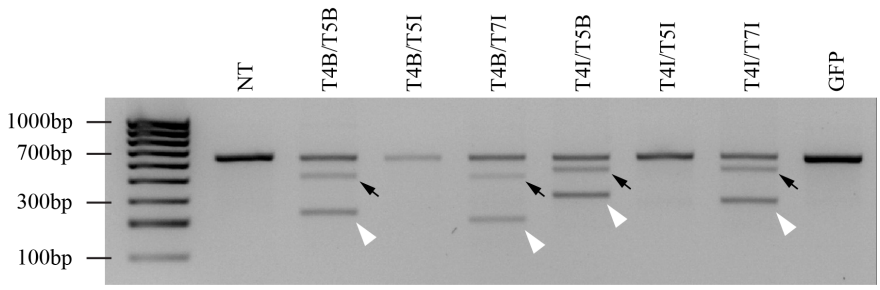


Figure 3-3. Co-transfection of HEK293T cells with combinations of CRISPR constructs can edit the C9orf72 gene. PCR amplification using a C9_OUT primer pair amplifies a non-edited band of ~650 bp in gDNA from treated and non-treated cells, but originates an additional smaller band (white arrowheads) in gDNA from HEK293T cells co-transfected for 48 hours with the different combinations of sgRNAs binding upstream and downstream of the repeat region. These bands correspond to the same amplified fragment but with partial deletions in the first intron, eliminating different sized fragments that include the hexanucleotide repeat region. Intermediate bands (black arrows) migrating between the non-edited and the edited bands are heteroduplex DNA products formed by hybridization of non-edited and edited DNA. NT = non-treated.

Analysis of these preliminary results already gave us some very relevant information. On one hand, it validated the basic concept behind our ideas, as the simultaneous expression of a pair of sgRNAs with SaCas9 was capable of introducing targeted deletions in the C9orf72 gene. On the other hand, we could exclude the T5I sgRNA from future experiments, as it shows very low activity when co-expressed with sgRNAs binding upstream of the G4C2 repeats (Figure 3-3).

3.3 Single-vector expression of C901 and C904 efficiently deletes C9orf72 targeted regions and promotes the re-ligation of the gene.

After the success of our co-expression paradigm, we sought to investigate whether expression of all therapeutic components from a single-vector construct would achieve similar results. Indeed, considering the ultimate aim of this project would be the translation of these strategies to the clinic using AAV-mediated gene therapies and considering the high costs of clinical-grade viral vector production (Ayuso, 2016), we initially hypothesised that the ability to use a single virus would be advantageous. Therefore, we used golden gate assembly to generate the pAAV-CRISPR-Cas9 single-vector expressing SaCas9 from a CMV promoter and two U6-driven sgRNA cloning sites (Figure 3-4). We then cloned the combinations of sgRNAs sequentially into this construct, after which we were able to do single transfections in HEK293T cells.

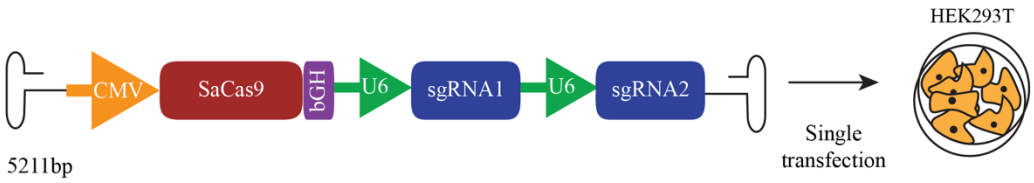


Figure 3-4. Single-vector construct expressing SaCas9 and two different sgRNAs. Schematic representation of the single pAAV vector cloned for this experiment, which expresses SaCas9 driven by the CMV promoter and two expression cassettes containing U6-driven cloning sites for insertion of sgRNAs followed by the sgRNA scaffold sequence. This vector can be transfected into HEK293T cells to introduce targeted deletions in the genome.

As per section 3.2, cells were kept for at least 48 h before harvesting for indel analysis. Interestingly, we obtained similar results as co-transfections, with PCR amplification of extracted gDNA using C9_OUT primers yielding edited bands (white arrowheads, Figure 3-5) in cells transfected with our single-vector construct cloned with the different combinations of sgRNAs. To further confirm the editing of the C9orf72, we TOPO-cloned the edited bands from two of the treated conditions (black arrows, Figure 3-5) and sent them for Sanger sequencing. Analysis of the sequencing confirmed that the human C9orf72 gene had been re-ligated across the predicted cut sites (determined as 3-4 bp before the PAM sequence), confirming the accuracy of this system and the excision of the targeted regions (sequencing traces, Figure 3-5) (Sanger sequencing of bands in **appendix 2**).

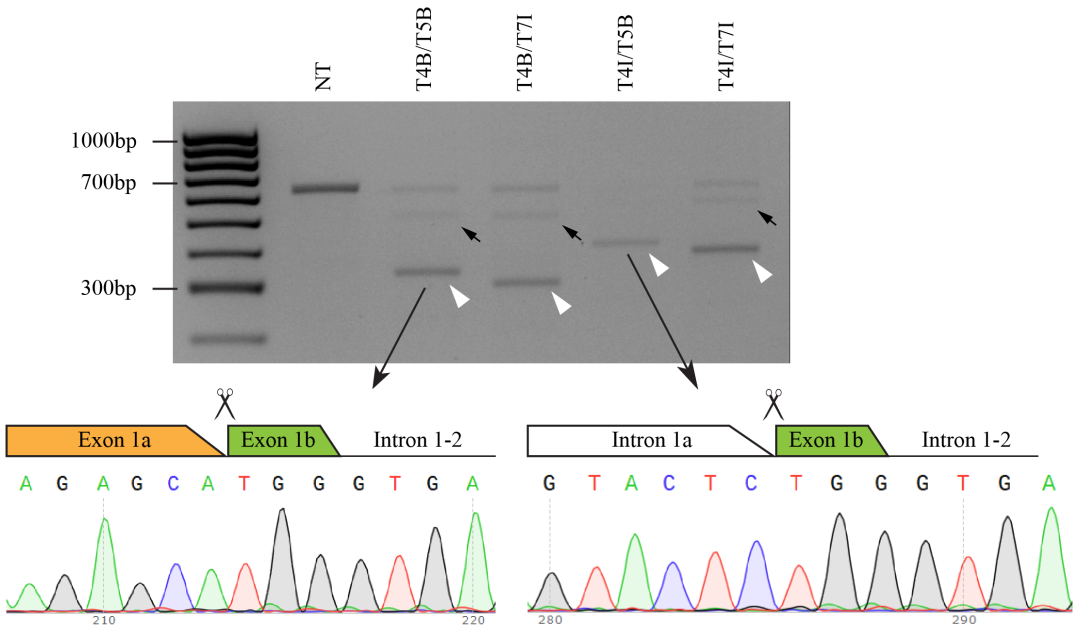


Figure 3-5. Indel screening of transfected HEK293T cells and sequencing confirmation of editing. PCR amplification using a C9_OUT primer pair amplifies a non-edited band of ~650 bp in gDNA from treated and non-treated cells. The same PCR originates an additional smaller band (white arrowheads) in gDNA from HEK293T cells transfected for 48 h with the single AAV CRISPR construct expressing different combinations of sgRNAs binding upstream and downstream of the repeat region. Intermediate bands (black arrows) migrating between the non-edited and the edited bands are heteroduplex DNA products formed by hybridization of non-edited and edited DNA. Edited bands from

two sgRNA combinations (long black arrows) were TOPO-cloned and sequenced (lower panels), aligning to the re-ligated C9orf72 gene around the predicted cut sites.

Upon confirmation that a single-vector delivery achieved comparable editing of the C9orf72, we then focused our efforts in reducing the size of this construct, to enable its packaging in an AAV-based viral vector. Indeed, AAV has a limited packaging capacity of approximately 4.7 Kb, with packaging over 5 Kb resulting in truncations of the transgene (Wu et al., 2010). To overcome this risk, we replaced the CMV promoter with the much smaller Elongation factor short 1a (EFS1a), which has been successfully used for viral gene delivery before (Tabebordbar et al., 2016). We then cloned a smaller, synthetic poly(A) signal (Gray et al., 2011; Levitt et al., 1989) to replace the bGH poly(A), further reducing the size of the construct to 4.65 Kb (Figure 3-6).

It was now warranted to check that these alterations to the promoter and poly(A) did not alter the vector's efficiency. To achieve this, we proceeded to the final selection of two pairs of sgRNAs that would fit our two different editing strategies: C901, which consists of excising the G4C2 repeats region; and C904, which targets and deletes the sequence immediately upstream of the repeats (Figure 1-7, chapter 1.5). From simple observation of the previous gel screenings (Figure 3-5), every pair of sgRNAs displayed would be a good candidate for the C901 therapy, but we selected the pair T4I/T5B (from now on, simply referred to as C901) for showing the highest intensity ratio of edited:unedited band. As for C904, the very short sequence length between the targeted Kozak consensus and the G4C2 repeats allowed for identification of only one good sgRNA candidate for binding downstream: T4D (black scissors, Figure 3-1). Considering the strong editing activity already shown with T4I, we selected the T4I/T4D pair of sgRNAs (from now on, simply referred to as C904) for the C904 therapy. We also designed a pair of sgRNAs, Ctrl1 and Ctrl2 (from now on, simply referred to as C902), that do not bind anywhere in the human or the mouse genome with up to four mismatches, to serve as editing controls. The abovementioned pairs of sgRNAs were then sequentially cloned into our shorter, modified AAV construct (Figure 3-6), after which we could transfect HEK293T cells for indel analysis.

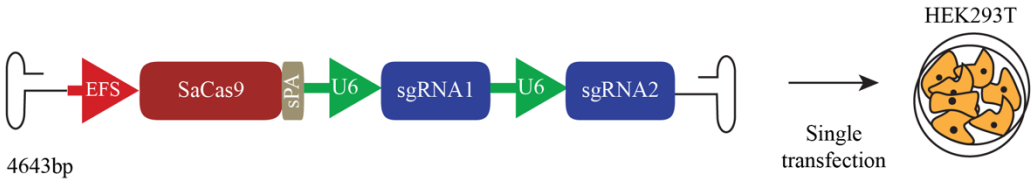


Figure 3-6. Short single-vector construct expressing SaCas9 and two U6-driven sgRNAs that can be inserted in an AAV virus. Schematic representation of the shorter version single-AAV vector cloned for this experiment, which expresses SaCas9 driven by the EFS1a promoter (~215 bp) and two expression cassettes containing U6-driven cloning sites for insertion of sgRNAs followed by the sgRNA scaffold sequence, flanked by AAV2 ITRs. Due to its size, this vector can be used to produce recombined AAV particles. The selected therapeutic sgRNA pairs T4I/T5B and T4I/T4D as well as the control pair Ctrl1/Ctrl2 were cloned sequentially in this vector to form therapeutic constructs C901, C904 and C902, respectively.

Upon single transfections of HEK293T cells for 48 h, PCR amplification with C9_OUT primer pairs from extracted gDNA revealed gene editing in cells transfected with plasmid C901 and C904 (pC901 and pC904, respectively), while all control conditions resulted in the sole amplification of the unedited version of the C9orf72 gene (Figure 3-7 A). The much smaller deletion introduced by C904 explains the bigger size of its edited DNA. Finally, we quantified the efficiency of precision editing and concluded that indeed this system elicited robust and reproducible editing of the C9orf72 in HEK293T cells, with approximately 45% efficiency for pC901 (Figure 3-7B) and 49% for pC904 (Figure 3-7 C).

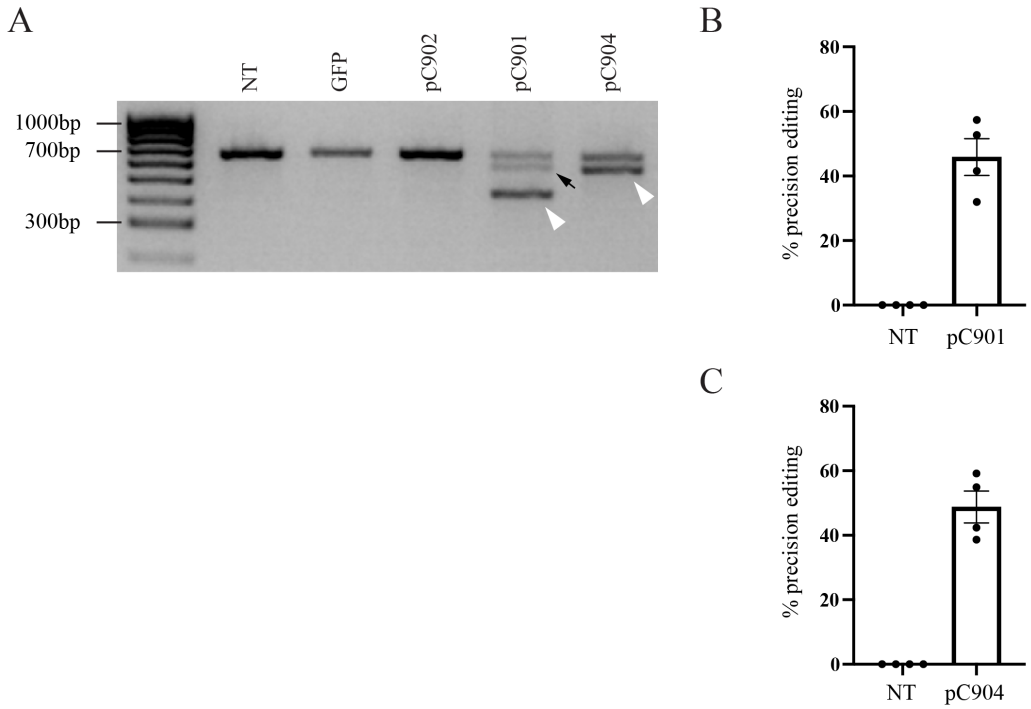


Figure 3-7. Transfection of HEK293T cells with the optimised single-AAV constructs induces robust and reproducible C9orf72 gene editing. (A) PCR amplification using a C9_OUT primer pair shows how transfection in HEK293T cells with pC901 and pC904 induces robust gene editing, as shown by amplification of smaller, edited bands (white arrowheads) in gDNA extracted 48 h post-transfection. The intermediate band (black arrow) migrating between the non-edited and the edited bands in the pC901 lane is the heteroduplex DNA product formed by hybridization of non-edited and edited DNA. (B and C) qPCR quantification estimates a 45% and 49% of precision editing by pC901 and pC904, respectively, upon 48 h transfections in HEK293T cells (n=4 biological replicates). Different methods were used to estimate pC901- and pC904-induced precision editing (see chapter 2.2.1.9 PCR-based methods for editing efficiency).

3.4 Editing the C9orf72 gene with C901 and C904 does not impact on the endogenous protein expression

One of the major concerns related to our therapeutic strategies was how the editing of the gene would impact on endogenous C9orf72 protein expression. In particular, considering the regions targeted by both strategies are in close proximity or even overlapping with the promoter regions (Haeusler et al., 2016), it was sensible to explore whether C901 or C904-mediated deletions could result in a decrease in C9orf72 protein levels. To test for this, we transfected HEK293T cells with pC901 and pC904 for 48 h, which we knew beforehand that originated robust gene editing and ran Western blots on protein extracts from these cells (Figure 3-8 A). We also transfected HEK293T with pC902 and with a plasmid encoding GFP, to serve as the editing and transfection controls, respectively. To make sure our C9orf72 antibody was specific, we used a HEK KO line created in-house using CRISPR to introduce a premature stop codon in exon 2 (Figure 3-8 A). Interestingly, quantification of these blots revealed that neither pC901 nor pC904 had any impact on WT C9orf72 expression (Figure 3-8 B). The absence of the band at the expected ~50 KDa in the KO line confirmed the specificity of the antibody used.

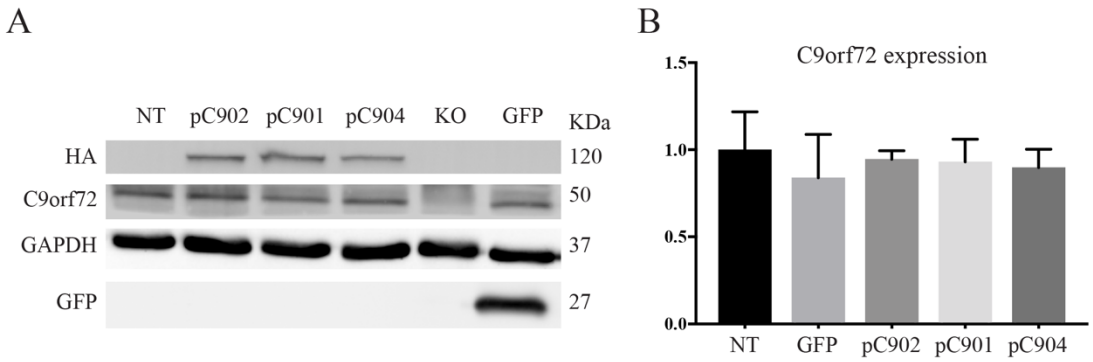


Figure 3-8. pC901 and pC904-mediated C9orf72 gene editing does not alter the expression of WT endogenous protein. (A) Western blot of extracts from HEK293T cells transfected with GFP, pC902, pC901 and pC904. GAPDH was used as loading control. (B) Western-

blot quantification shows that neither pC901 nor pC904 significantly affect the WT C9orf72 protein expression 48 h after transfection (n=3 biological replicates). One-way ANOVA, ns = not significant.

3.5 Discussion

In this chapter we describe the conceptualization of the project, the design of our proposed therapeutic strategies, their initial *in vitro* validation in cells without a pathogenic C9orf72 expansion and the optimisations of the vector into a gene therapy construct ready to be packaged into AAV viruses. The design of the single-vector construct was one of the most technically challenging aspects of this stage due to the large size of the transgene, requiring creative solutions and forcing us to take several considerations. For instance, even though SaCas9 cDNA (~3.2 Kb) is considerably smaller than the most commonly used SpCas9 cDNA (~4.2 Kb), its size immediately rules out using self-complementary AAV (scAAV). Having been used with great success in the clinic to treat SMA (Mendell et al., 2017), scAAV is an attractive configuration because it mediates earlier and more efficient transgene expression than single-stranded AAV (ssAAV) (McCarty et al., 2001; Wang et al., 2003). However, because of the transgene size, we were limited to cloning our constructs into a ssAAV vector. In addition to this, our editing strategy requires expression of two sgRNAs, bringing the size of the transgene to approximately 5.2 Kb (Figure 3-4), going over the packaging limit of AAV and risking 5' viral genome truncations.

One potential solution to reduce the size of this vector would be to express both sgRNAs from a single promoter, using a polycistronic gene that can be transcribed and processed into two separate sgRNAs. Several mechanisms have been designed so far that could achieve this, the first of them consisting in expressing both sgRNAs linked by a sequence recognizable by the Csy4 RNA nuclease from *Pseudomonas aeruginosa*, co-expressed by the same construct, which then cleaves the target sequence and releases the two sgRNAs (Nissim et al., 2014; Tsai et al., 2014). Importantly, this system allows to express the sgRNAs from either the commonly used, ubiquitous RNA polymerase III U6 promoter, or from an RNA polymerase II promoter, allowing tissue specific expression. Even though this creative mechanism can be very useful

for many CRISPR-mediated multiplexing experiments, its usefulness increases with the number of sgRNAs to be expressed. Since our strategy requires only two, and considering the need to co-express the Csy4 nuclease, we would still need to provide the Cas9 from a separate construct, effectively neutralizing its advantage in this project. A better alternative for our strategies would consist in using ribozymes, a class of enzymatic, self-cleaving RNA molecules (Kruger et al., 1982), that were first used to circumvent the need for RNA polymerase III promoters to express sgRNAs in yeast (Gao and Zhao, 2014) and *Leishmania donovani* (Zhang and Matlashewski, 2015). We could use this strategy to flank each sgRNA with the self-cleaving ribozymes Hammerhead (HH) and Hepatitis delta virus (HDV) on its 5' and 3' end, respectively, as was reported previously to successfully edit human cells (Nissim et al., 2014; Yoshioka et al., 2015). This would bypass the need to use a U6 promoter, as we could drive expression of every component in our transgene cassette from the CMV promoter which, considering the small size of these ribozymes, would reduce the size of the construct (Figure 3-9 A). Alternatively, we could also express both gRNAs from a single U6 promoter if we linked both guides with the HDV and HH ribozymes (Xu et al., 2017) (Figure 3-9 B). If, however, we wanted to avoid the use of ribozymes altogether, there is another elegant approach that would be compatible with our therapeutic strategies. This approach takes advantage of the highly conserved tRNA processing system, in which pre-tRNAs are cleaved at specific sites to trim its 5' and 3' extra sequences. This system was first demonstrated by Xie and colleagues (Xie et al., 2015) who, by linking two or more sgRNAs with specific tRNA sequences, were able to use the endogenous cell machinery to process a polycistronic transcript in rice plants. Several adaptations of this system were successful in other plant species (Hui et al., 2019; Wang et al., 2018) and also in human cells (Dong et al., 2017), which led us to consider adapting this strategy in our construct by linking our sgRNAs with tRNA or combine its use with ribozymes (Figure 3-9 C and D). However, a study comparing the ribozymes and tRNA polycistronic systems in parallel found that these are slightly less efficient in their induced editing than expressing each sgRNA under its own U6 promoter, due to imperfect cleavage of the two sgRNAs (Xu et al., 2017), for which reason we ultimately decided to go with this conformation (Figure 3-4).

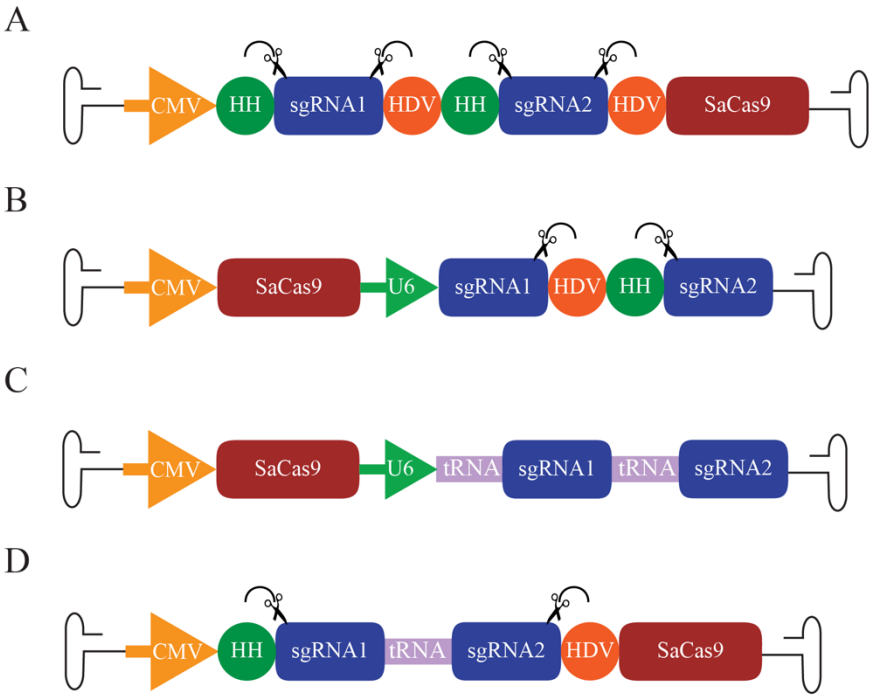


Figure 3-9. Alternative conformations for the single-vector construct used in this project. (A and B) Alternative strategies to reduce the size of our therapeutic single-vector construct using self-cleaving HH and HDV ribozymes, allowing to drive expression of the entire construct from the CMV without the need for a polymerase III promoter or (B) expressing both sgRNAs from a single U6 promoter. (C) A different conformation which takes advantage of the cellular endogenous tRNA cleaving system by using specific tRNA linker sequences to separate both sgRNAs and drive their expression from a single U6 promoter. (D) A combination of the ribozymes and tRNA strategies, allowing to express the entire transgene from the CMV and saving space that would otherwise be occupied by the U6 promoters. These concepts were shown to work in one way or another with spCas9 in a number of projects spanning several animal and plant models (Dong et al., 2017; Gao and Zhao, 2014; Hui et al., 2019; Nissim et al., 2014; Wang et al., 2018; Xie et al., 2015; Xu et al., 2017; Yoshioka et al., 2015; Zhang and Matlashewski, 2015).

With the option of a polycistronic gene excluded for now, we decided to reduce the size of the single vector by replacing the CMV promoter with a smaller one, such as the EFS1a (Kim et al., 1990; Nelson et al., 2010) or MeP (Gray et al., 2011), with approximately 215 bp and 230 bp, respectively. The MeP promoter was reported to induce low, long-term expression in the CNS and was even used in an SpCas9 construct to mediate *in vivo* gene editing (Swiech et al., 2014). However, the widespread and well-described use of EFS1a in gene editing applications driving either SpCas9 or SaCas9 from AAV and lentiviruses in various tissue and cell types, both *in vitro* and *in vivo* gave us confidence to adapt it for this project (Escobar-Aguirre et al., 2019; Heckl et al., 2014; Kumar et al., 2018; Nishiyama et al., 2017; Tabebordbar et al., 2016; Vora et al.). Nevertheless, we have still generated the constructs in Figure 3-6 with the MeP promoter as a contingency plan, and we have future experiments planned to investigate whether it mediates stronger editing *in vitro* and *in vivo* compared with the EFS1a. Replacing the promoter reduced the size of the construct to approximately 4.85 Kb, which can fit into an AAV particle, although packaging transgenes over 4.7 Kb can result in reduced viral titer. To avoid this, we cloned in a minimal 49 bp synthetic poly(A) signal that has been used before to package large transgenes (Gray et al., 2011; Levitt et al., 1989; Swiech et al., 2014), although its use can result in reduced transgene expression when compared with the bGH poly(A) (Azzoni et al., 2007).

In 2017, Kim and colleagues (Kim et al., 2017) reported an even smaller Cas9 nuclease from *Campylobacter jejuni*, CjCas9, whose cDNA is approximately 200 bp smaller than SaCas9, which makes it a very attractive alternative for *in vivo* gene editing projects. Although the timing of its publication was incompatible with this project, it is interesting to speculate whether we can use CjCas9 for our editing strategies in the future since it would allow us to use either a slightly bigger promoter than the EFS1a or, alternatively, use the bGH poly(A) signal.

As discussed above, all the changes made to the construct by us were the result of careful consideration and reaching a compromise between transgene expression strength and error-free AAV packaging efficiency. Interestingly, these modifications did not impact the final vector (Figure 3-6) negatively, as evidenced by its high efficiency in editing the C9orf72 gene in HEK293T cells (Figure 3-7 B and C). Indeed, this efficiency might be even higher than the one calculated, since the qPCR assay used is likely to underestimate the editing rate of the C901 therapeutic strategy (see chapter 2.2.1.9 PCR-based methods for editing efficiency).

The last experiments of this chapter focused on investigating the outcome of editing the C9orf72 gene on the endogenous protein production. Indeed, the design of our experimental strategies implicates that the WT allele of the gene will be edited just like the mutant, expanded allele. Considering that the vast majority of C9orf72 patients are heterozygous, it was important to assess the effect of our gene editing therapies in a normal, WT allele without a pathogenic expansion. Although all regions targeted by our sgRNAs are non-coding, they either overlap or locate in near proximity of the promoter regions (Haeusler et al., 2016), for which reason we feared mutating those sequences. Also, despite contradicting reports in the literature, haploinsufficiency is still commonly proposed as a mechanism to explain expanded C9orf72-mediated neurodegeneration (Shao et al., 2019; Shi et al., 2018), and as such, decreasing WT protein expression would not make for an attractive therapeutic option. The experiments we describe, however, conclusively prove that even with ~50% of every WT C9orf72 allele being edited by C901 and C904 in a given cell population, the endogenous protein levels remain unchanged.

To summarise, in this chapter we have described the design of two new experimental therapies based on a highly targeted CRISPR-mediated editing of the C9orf72 gene. We have developed an ITR-flanked AAV construct that can be cloned with any pair of sgRNAs to introduce genomic deletions or chromosomal translocations via AAV delivery. We have cloned this construct with the appropriate sgRNAs and demonstrated how it can induce precise, programmable deletions in the C9orf72 gene of a human cell line. Finally, we have demonstrated how robust C901 and C904-mediated editing does not alter the WT protein expression. The next step will be to produce AAV viruses and testing these experimental therapies in cells with a C9orf72 gene containing a hexanucleotide repeat expansion.

4. *In vitro* testing: cells with expanded C9orf72

4.1 Introduction

In the previous chapter, we have described the rationale behind our therapeutic approaches, the design of our strategies and their components, and an *in vitro* proof-of-concept in HEK293T cells that have 3xG4C2 repeats. Although they showed promise, it was difficult to predict how these experiments would translate into clinically relevant models, owing mainly to two different difficulties that we could foresee. The first issue concerns the delivery of the therapeutic constructs into the target cells, as the remarkably high transfection rates routinely achieved in HEK293T cells are unlikely to be replicated in any other cell types. Second, the presence of the hexanucleotide expansion with its instability, highly mutagenic nature and strong secondary structures could pose difficulties to the Cas9-mediated editing. It was then warranted to test our therapeutic approaches in more relevant disease models, particularly with a mutated C9orf72 gene.

The C9-500 BAC mouse model described by Liu and colleagues (Liu et al., 2016) was created by inserting an entire copy of a human C9orf72 gene with 500xG4C2 repeats including introns and substantial portions of flanking sequence, originally derived from a patient, into the mouse genome of an FVB/NJ background. In their study, the authors described how a sub-population of the C9-500 BAC transgenic animals develop a fast-progressing disease with C9orf72-ALS hallmarks such as progressive paralysis and decreased survival, in addition to neurodegeneration, accumulation of RNA Foci and DPRs, and NMJ denervation. Furthermore, the characterised animals are hemizygous, meaning they lack a WT copy of human C9orf72. Since the mouse orthologue of C9orf72 differs substantially from the human gene in its 5' region, this means that any gene editing events near the G4C2 repeat region in the hemizygous' genome and its downstream consequences would be linked exclusively to the expanded allele. In contrast, testing our experimental therapies in patient-derived human cells would pose a significant problem when trying to assess its editing efficacy in the presence of a C9orf72 expansion, since we wouldn't be able to trace the origin of edited DNA to the mutant or the WT allele.

Considering these reasons, the C9-500 BAC model presented itself as an ideal platform to start testing our therapeutic strategies in disease-relevant cells after their initial validation in the HEK293T cell line. In this chapter, we describe our first steps in exploring the suitability of

primary cortical neurons cultured from these mice as an initial screening platform for our proposed therapies. We then report how we took advantage of this quick and easy access to C9-500 BAC neuronal cells to test our editing strategies, namely their ability to edit mutated human C9orf72 and its impact on pathophysiological hallmarks of disease like RNA Foci and DPRs. We expected that these experiments would give us an early indication of the therapeutic potential of our strategies and inform us on whether to start testing these *in vivo*.

4.2 Validating cortical neurons cultured from C9-500 BAC mice as an *in vitro* model to test C9orf72 therapeutic editing

As an initial *in vitro* platform to start testing our therapeutic strategies in cells with a human expanded C9orf72 gene, we decided to use primary cortical neurons cultured from C9-500 BAC mouse embryos. Considering that the C9orf72 copy used to create this model was cloned from a human patient (Liu et al., 2016), we wondered whether any patient-specific mutations in the genomic region flanking the expansion would create mismatches between the designed sgRNAs and their target sequence. To explore this possibility, we decided to PCR-amplify and sequence the regions flanking the repeat expansion in DNA extracted from these mice. We started by designing primers to amplify a ~180 bp sequence upstream of the G4C2 repeats (Figure 4-1 A), whose PCR yielded a product at the predicted size (Figure 4-1 B). TOPO-cloning of the band followed by Sanger sequencing showed that the region upstream of the expansion in this model aligns perfectly to the human C9orf72 sequence retrieved from ENSEMBL. Indeed, close analysis of the sequencing data reveals no genetic mutations in the DNA targeted sequence of the T4I and T4D sgRNAs used in this project (Figure 4-1 C) (Sanger sequencing of band in **appendix 3**).

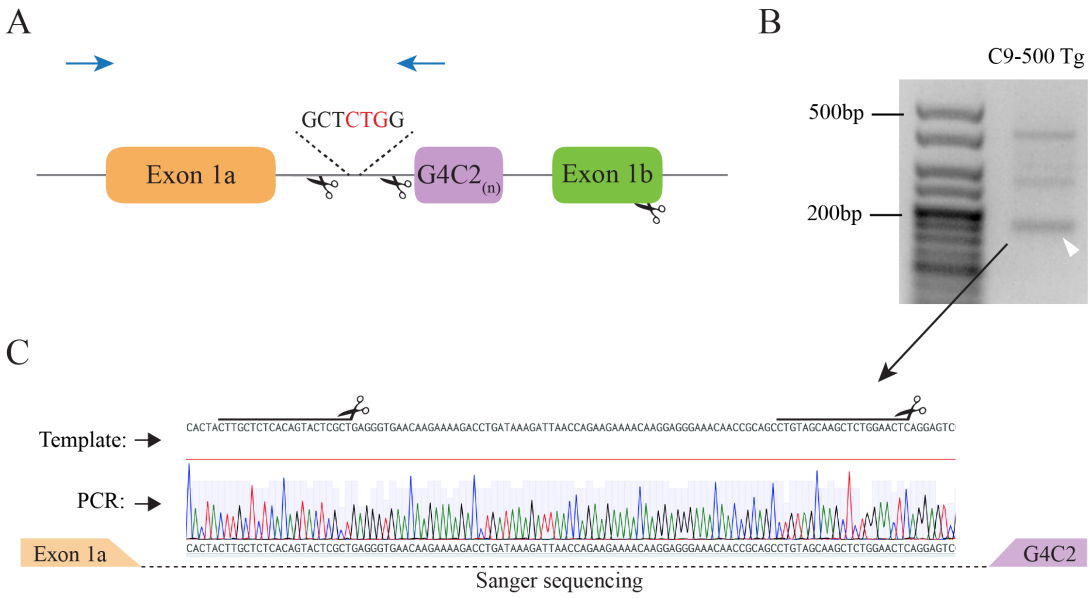


Figure 4-1. Investigation of potential mutations in the region upstream of the G4C2 repeats in C9-500 BAC Tg mice. (A) Schematic representation of the PCR designed to amplify the region directly upstream of the repeat region, spanning exon 1a and the first half of intron 1. The primer pair C9_INT_1 was designed for this PCR, represented by blue arrows. (B) Agarose gel of the PCR reaction shows successful product amplification at the expected size (white arrowhead). (C) Alignment of Sanger sequencing traces of TOPO-cloned PCR band reveals perfect complementarity with the human C9orf72 DNA template used to design the sgRNAs. Black lines with scissors represent sgRNA binding sites, showing no mutations that could prevent the correct binding of the guides. The alignment panel was cropped from Benchling, the software used to align the sequences (Benchling, 2020).

Likewise, we designed PCR primers to amplify a ~840 bp region located directly downstream of the expansion in DNA from the C9-500 BAC model (Figure 4-2 A and B). Interestingly, TOPO-cloning of the amplified DNA followed by Sanger sequencing also revealed perfect complementarity of the patient DNA used to generate the C9-500 BAC model to the reference human C9orf72 gene, in the T5B sgRNA targeted region (Figure 4-2 C) (Sanger sequencing of band in **appendix 3**). These results suggest that the sgRNAs used in C901 and C904 editing strategies should bind to their targeted C9-500 BAC sequence without any mismatches.

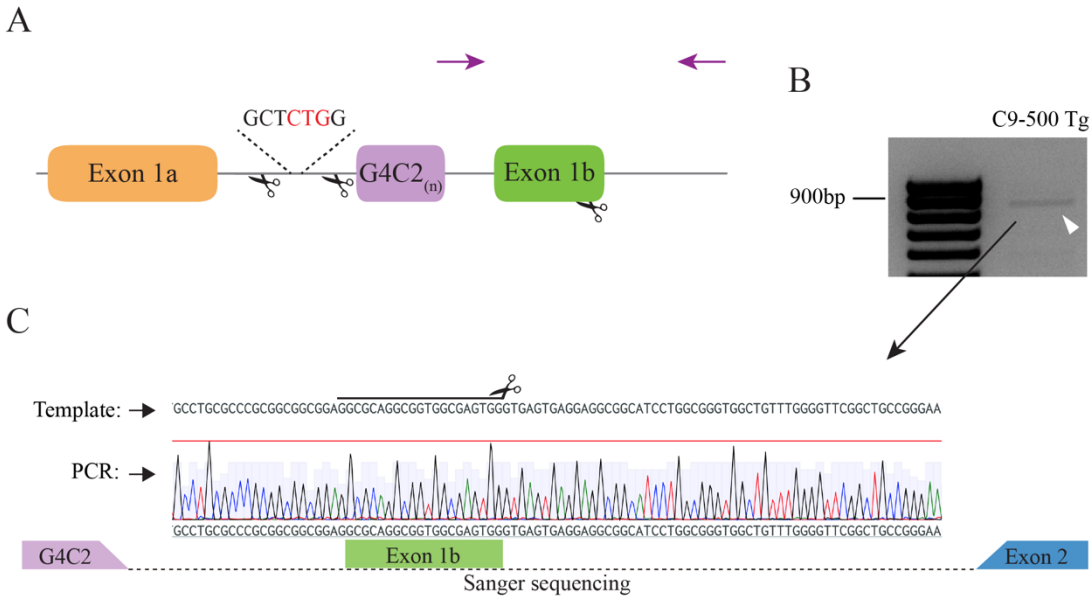


Figure 4-2. Investigation of potential mutations in the region downstream of the G4C2 repeats in C9-500 BAC Tg mice. (A) Schematic representation of the PCR designed to amplify the region directly downstream of the repeat region, spanning the non-coding exon 1b and several bp of intronic sequence. The primer pair C9_EXT_3 was designed for this PCR, represented by purple arrows. (B) Agarose gel of the PCR reaction shows successful product amplification at the expected size (white arrowhead). (C) Alignment of Sanger sequencing traces of TOPO-cloned PCR band reveals perfect complementarity with the human C9orf72 DNA template used to design the sgRNAs. Black line with scissors represents T5B sgRNA binding site, expressed by the C901 therapy construct, showing no mutations that could create potential mismatches. The alignment panel was cropped from Benchling, the software used to align the sequences (Benchling, 2020).

Having confirmed sequence complementarity of the model with the sgRNAs used in this project, we were ready to start testing our editing strategies in cortical neurons from C9-500 BAC mice. Primary neuronal cultures are considerably difficult to transfect chemically (Karra and Dahm, 2010; Sariyer, 2013), for which reason we decided to use viruses, particularly AAV9, which are highly efficient delivery vehicles to postmitotic cells. In addition to this,

AAV9 has an excellent safety profile and has been used with remarkable success in pre-clinical and clinical trials to treat neurological diseases, making it an ideal viral vector for potential clinical applications (Iannitti et al., 2018; Mendell et al., 2017; Meyer et al., 2015). However, before producing the therapeutic viruses, it was warranted to explore the ability of AAV9 to transduce cortical neurons from the C9-500 BAC model and the suitability of the EFS1a promoter for this cell type. To check for this, we transduced cortical neurons from non-Tg embryos with an AAV9 virus expressing GFP under the EFS1a promoter and observed high transduction levels, with over 55% of GFP-expressing neurons, demonstrating the suitability of the AAV9 capsid and the promoter for these experiments (Figure 4-3).

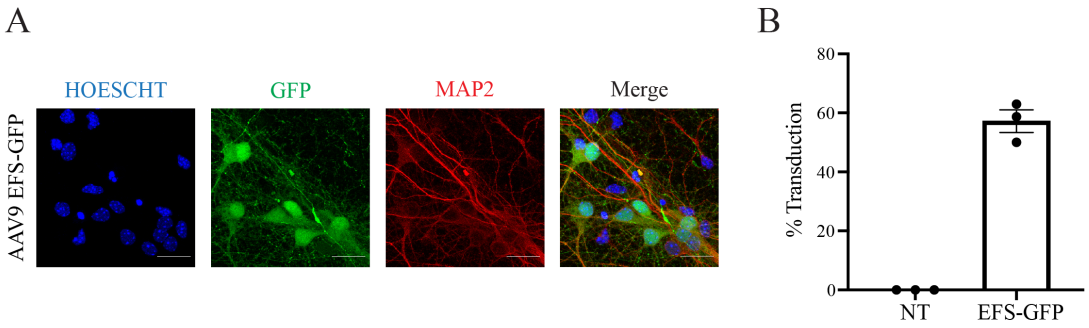


Figure 4-3. The AAV9/EFS1a promoter system mediates efficient gene transfer in cultured cortical neurons from C9-500 BAC mice. (A) Confocal microscope images showing how 6-day transduction of cortical neurons cultured from C9-500 BAC mice with AAV9 EFS1a-GFP induces efficient GFP expression, demonstrating that the capsid and the promoter can be used to mediate transgene expression in these cells. Scale bar = 20 µm. (B) Quantification of GFP-positive neurons transduced with AAV9 EFS1a-GFP for 6 days, showing over 55% transduction efficiency (n=3 biological replicates).

Taking these results into account, we proceeded to make large-scale productions of recombinant AAV9 viruses from all our CRISPR constructs: C901, C904 and our control sgRNA construct C902. Western blot and qPCR confirmed that all viral productions successfully assembled intact AAV9 particles and with high viral titers, respectively (Figure 4-4) (Table 4.1).

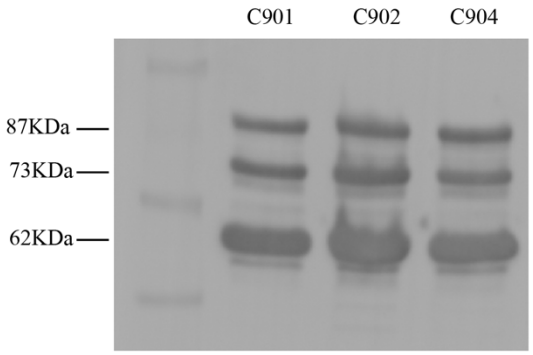


Figure 4-4. Viral production of C901, C902 and C904 yields intact AAV9 particles. Western blot of viral extracts from C901, C902 and C904 viruses probed with an antibody against AAV capsid proteins VP1 (87 KDa), VP2 (73 KDa) and VP3 (62 KDa) shows the corresponding bands at the correct sizes, demonstrating the capsid integrity of the viruses produced.

Table 4.1. AAV9 viral titers. List of titers of AAV9 productions used in this chapter, expressed as viral genome copies per ml (vg/ml).

Virus	**Titer**
AAV9 C901	$4.48x10^{13}$ vg/ml
AAV9 C901 (2)	$7.14x10^{13}$ vg/ml
AAV9 C902	$1x10^{13}$ vg/ml
AAV9 C902 (2)	$5.72x10^{13}$ vg/ml
AAV9 C904	$2.83x10^{13}$ vg/ml
AAV9 C904-CHOP	$1x10^{13}$ vg/ml

Having produced the viruses, it was important to confirm whether these viral vectors were able to express SaCas9 in the C9-500 BAC cortical neurons. With this objective in mind, we transduced cortical neurons from non-Tg littermates with our different recombinant AAVs, to carry out immunocytochemistry. Indeed, in all of our constructs, SaCas9 is tagged with an HA epitope which allows the tracking of transduced cells. This study revealed that all of the produced viruses were able to mediate expression of HA-tagged SaCas9 (Figure 4-5 A) in these neurons with an efficiency ranging from approximately 15 to 30% (Figure 4-5 B). There was no statistical difference in the transduction efficiency between the different viral treatments.

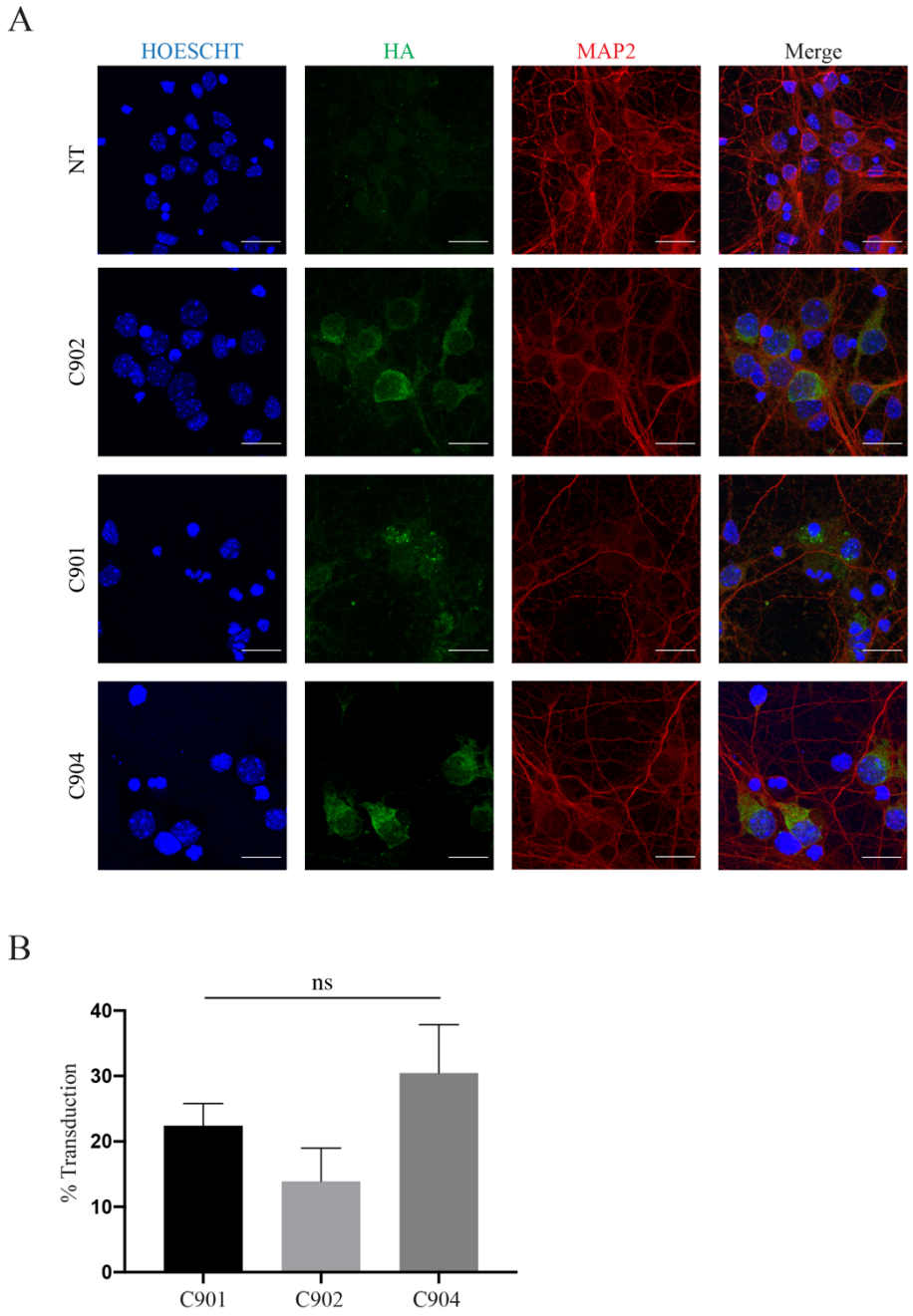


Figure 4-5. AAV9 transduction induces expression of HA-tagged Cas9 in cortical neurons from C9-500 BAC mice. (A) Confocal microscope images showing HA expression upon 6-day transduction of C9-500 cortical neurons with C901, C902 or C904 viruses (600'000 vg/cell). Scale bar = 20 µm. (B) Quantification of HA-positive neurons transduced with C901, C902 or C904 for 6 days, showing no significant differences between treated groups (n=3 biological replicates). One-way ANOVA, ns = not significant.

Since the ultimate aim of these experiments was to evaluate therapeutic efficacy of editing the C9orf72, we decided to investigate whether C9-500 BAC cortical neurons displayed pathological hallmarks of disease, such as accumulation of repeats-containing RNA Foci. Being direct by-products of expanded C9orf72, and considering that each Foci is composed of a single mutant transcript (Liu et al., 2017), expression of RNA Foci could be used as a sensitive method to evaluate immediate downstream effects of C9orf72 gene editing events. Also, there is evidence to support their role as an important disease mechanism (Česnik et al., 2019; Cooper-Knock et al., 2014a; Zu et al., 2013), for which reason they could be used as an early indicator of therapeutic potential. Considering this, we cultured cortical neurons from C9-500 BAC transgenic and non-Tg embryos and carried out FISH staining, which confirmed that these neurons display high levels of sense RNA Foci (Figure 4-6), located mainly in the nucleus. RNAse A-treated neurons did not show any expression of Foci, demonstrating the specificity of this assay.

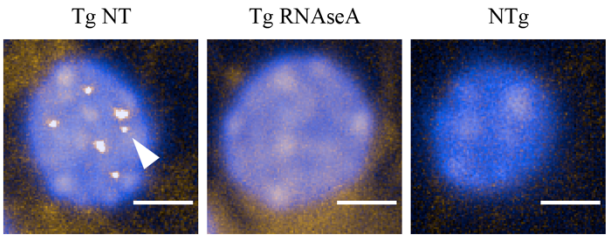


Figure 4-6. Cortical neurons cultured from C9-500 BAC mice display sense RNA Foci in the nucleus. Confocal microscope (Opera Phenix, Perkin Elmer) images showing the presence of sense nuclear (white arrowhead) RNA Foci in cortical neurons from transgenic C9-500 BAC embryos. These Foci are absent in RNAse A-treated Tg neurons and in cortical neurons from non-Tg littermate embryos. Scale bar = 5 µm.

These results put together demonstrate how cortical neurons cultured from the C9-500 BAC mice constitute an excellent *in vitro* model to study gene modulation of C9orf72 for therapeutic purposes. Indeed, it provides a quick access to primary neuronal cells containing a pathogenic human C9orf72 allele, whose G4C2 flanking regions are conserved and mutation-free. In addition, these cells can be transduced with the clinically relevant AAV9 virus and display nuclear RNA Foci, which can be used as a screening method for therapeutic potential.

4.3 Transduction with C901 and C904 successfully edits human C9orf72 in cortical neurons and decreases key expansion by-products

Having established the suitability of primary cortical neurons from C9-500 BAC model to test our therapeutic strategies, we proceeded to check whether we could edit a human C9orf72 gene with a pathogenic G4C2 expansion. The genetic design of this model makes it ideal for these experiments, since hemizygous mice do not have a WT allele and the mouse orthologue of C9orf72 differs from the human sequence in its 5' region. Considering this, we are able to pinpoint any editing events to the human pathogenic allele exclusively. We started by culturing cortical neurons from 3 different transgenic embryos for 8 days, which we transduced with our different viruses on day 2. Even though they were genotyped as transgenic, we took the additional step of confirming the presence of a G4C2 pathogenic expansion by analysing DNA from each of the 3 embryos by repeat-primed PCR (Figure 4-7).

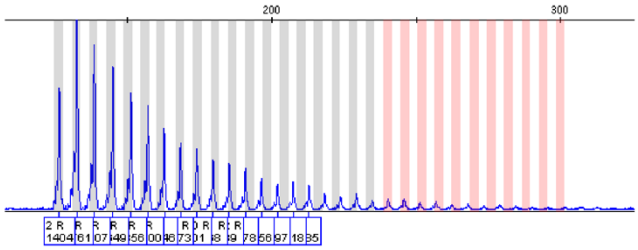


Figure 4-7. Repeat-primed PCR confirms G4C2 expansion in cortical neurons from used embryos. DNA extracted from each of the 3 embryos used in this experiment was analysed by repeat-primed PCR and tested positive for the presence of the hexanucleotide expansion. This image represents the traces from one of these embryos. The other traces can be found in **appendix 4**.

At the end of the experiment, DNA extracted from treated neurons was analysed by endpoint PCR across the 5' region of the C9orf72 gene to check for precision editing. This PCR did not originate a product in non-treated or C902-treated cells, presumably due to the 100% GC nature of the repeat expansion (Figure 4-8 A), as was reported previously (DeJesus-Hernandez et al., 2011). However, in neurons transduced with C901 for just 6 days, the deletion of the entire repeat region allowed the polymerase to read through and originate a product (arrowheads, Figure 4-8 A). TOPO-cloning of the edited bands followed by Sanger sequencing confirmed that the C9orf72 gene had been edited and re-ligated with remarkable precision across the predicted cut sites for sgRNAs T4I and T5B, having essentially introduced a deletion in the region that included the G4C2 repeats (Figure 4-8 B and C). Quantification by qPCR of all 3 samples estimated a ~1.5% rate of precision editing induced by C901 in C9-500 BAC neurons.

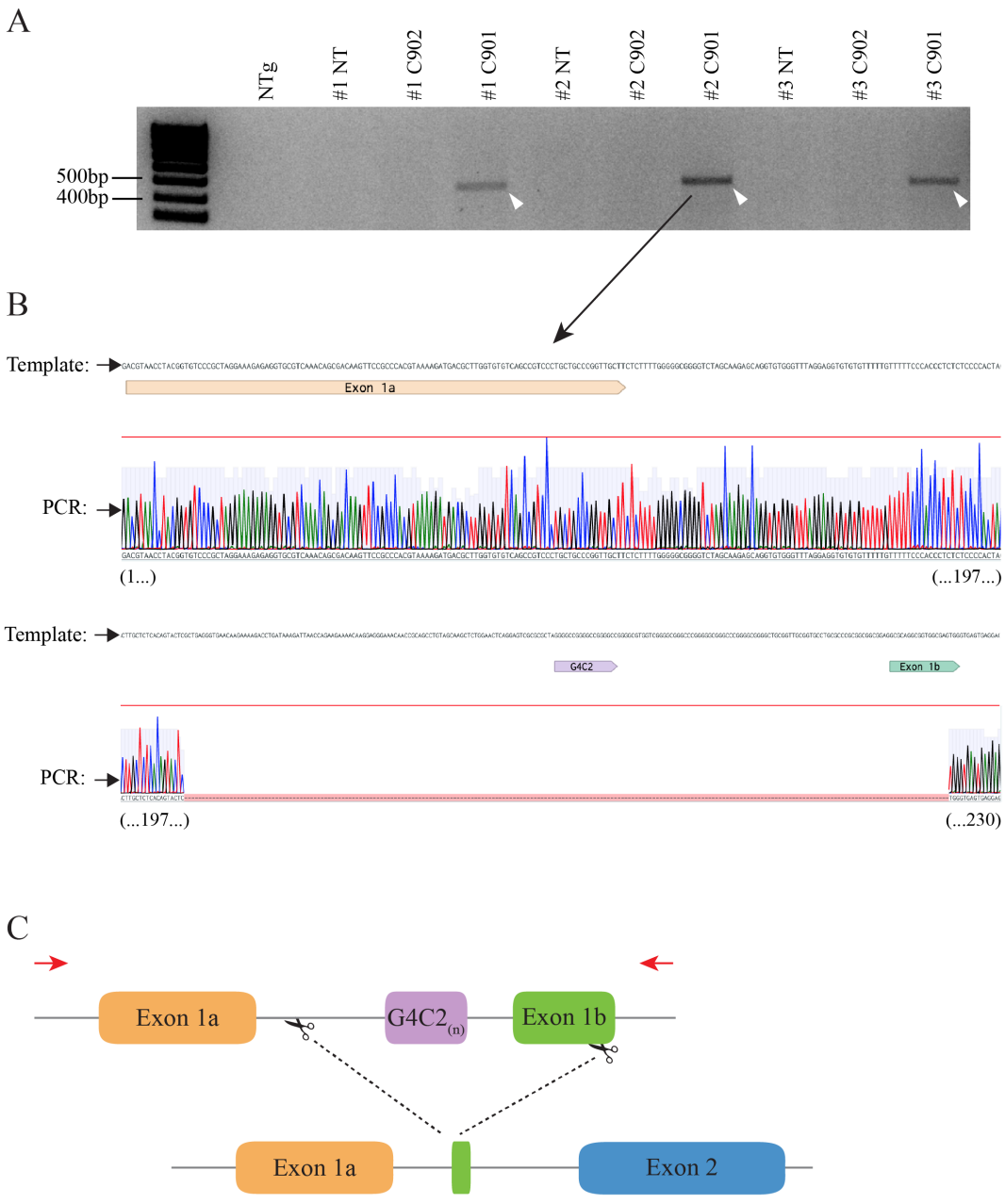


Figure 4-8. C901 can delete a full-size G4C2 expansion from the human C9orf72 gene. (A) Agarose gel showing consistent excision of a full-size C9orf72 pathogenic expansion with C901, in cortical neurons derived from the C9-500 BAC mouse model. PCR amplification with C9_OUT_2 primers flanking the GC-rich repeats does not give any product in non-treated (NT) or C902-transduced neurons, since the polymerase cannot read through the expansion. In neurons transduced with a high dose of C901 (600'000 vg/cell), the G4C2 expansion is excised enabling the polymerase to read through the sequence and originate a product (white arrowheads). #1, #2 and #3 represent cortical neurons

cultured from 3 different embryos. (B) Alignment of Sanger sequencing traces of a TOPO-cloned edited band. The lower panel shows how the edited DNA aligns to the human C9orf72 gene excluding the portion cut by C901, demonstrating that the gene re-ligated across the predicted cut sites. The panels are cropped images of alignments using Benchling (Benchling, 2020). (C) Schematic representation of the C901 strategy and its editing on the C9orf72 region of interest. PCR amplification using C9_OUT_2 primers (red arrows) in cortical neurons from C9-500 BAC mice does not amplify any product, due to secondary structures and the GC-rich nature of the expansion, and the absence of a WT allele. C901-mediated cutting promotes the excision of the repeats and re-ligation of the C9orf72 gene, allowing the PCR to yield a product.

Following a similar strategy to the C901 screening, we proceeded to confirm whether the C904 strategy was also able to edit an expanded human C9orf72 allele, considering the close proximity of its target sites to the hexanucleotide expansion. Therefore, we cultured C9-500 BAC cortical neurons for 8 days, including a 6-day transduction period with a high dose of C904 virus. Upon genomic DNA extraction, a PCR was run using a pair of primers designed to amplify the region immediately upstream of the G4C2 repeats. This PCR yielded a ~400 bp band in DNA from all transgenic neurons, but also amplified an additional, slightly smaller product corresponding to the exact same portion of the C9orf72 gene with a targeted ~90 bp deletion (Figure 4-9 A). To confirm our hypothesis, we TOPO-cloned the edited bands upon extracting the DNA from the agarose (arrowheads, Figure 4-9 A) and sent the samples for Sanger sequencing. Alignment to human C9orf72 and analysis of the traces confirms the deletion introduced in the region upstream of the G4C2 repeats, effectively removing the targeted Kozak sequence from the gene (Figure 4-9 B and C). By calculating the band intensity ratios we estimated a ~14% precision editing induced by C904 in these neurons across 3 different transgenic embryos, a considerably higher editing rate than that of C901 virus in these cells. However, it is important to note that the apparent improvement in precision editing efficiency of C904 compared with C901 might be linked to the different methods used to calculate precision editing, and not necessarily due to C904 being a more efficient editing strategy.

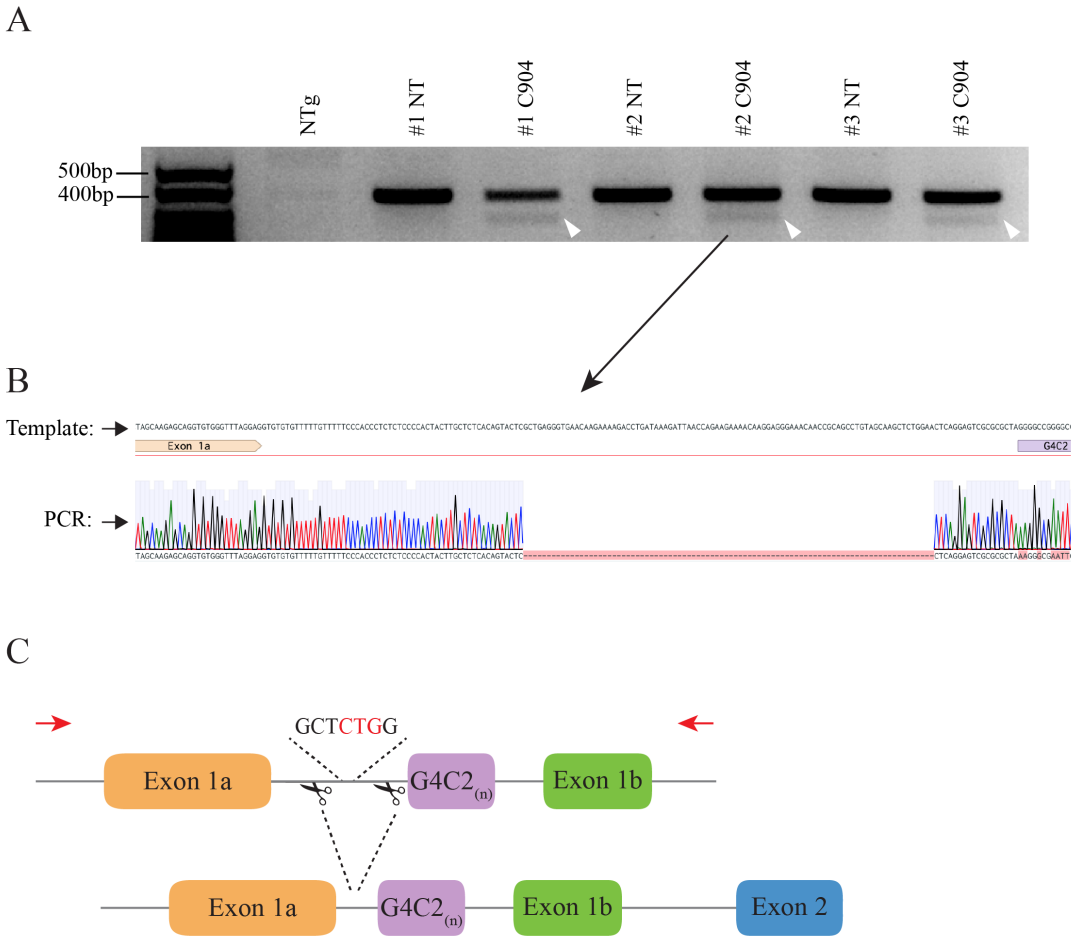


Figure 4-9. C904 can introduce a targeted deletion in the regulatory region upstream of the G4C2 expansion in human C9orf72. (A) Agarose gel showing C904-mediated editing of the C9orf72 gene in C9-500 cortical neurons upon PCR amplification with C904_INT_14 primers. Transduction with a high dose of C904 therapeutic virus (600'000 vg/cell) introduces a targeted ~90 bp deletion originating edited bands at the expected size (white arrowheads) just below the unedited product. #1, #2 and #3 represent cortical neurons cultured from 3 different embryos. (B) Alignment of Sanger sequencing traces of a TOPO-cloned edited band. The panel shows how the edited DNA aligns to the human C9orf72 gene excluding the portion excised by C904, a ~90 bp fragment that includes the targeted Kozak sequence. The panel is a cropped image of an alignment using Benchling software (Benchling, 2020). (C) Schematic representation of the C904 strategy and its editing of the first intron of the C9orf72. PCR amplification using C904_INT_14 primers (red arrows) in cortical neurons from C9-500 BAC mice amplifies a product of ~400 bp. The same PCR done in DNA extracted from C904-treated cells amplifies an additional, ~90 bp smaller product, resulting from the deletion strategically introduced in the regulatory sequence upstream of the repeat expansion.

Having established the ability of our therapeutic viruses to edit expanded C9orf72 in primary neurons, it was now important to investigate whether this editing had an impact in the key pathological hallmarks of C9orf72, namely RNA Foci and DPR production and accumulation. Deriving from the direct transcription of the G4C2 repeats, RNA Foci constitute the most upstream by-product of the C9orf72 expansion, for which reason we focused on this readout first. Having previously established the presence of Foci in C9-500 BAC transgenic cortical neurons (Figure 4-6), we transduced these cells with our different AAV CRISPR viruses for 6 days and scanned the plates with the high-throughput confocal microscope Opera Phenix (Perkin Elmer) after FISH staining. Automated, unbiased quantification of FISH-positive nuclear inclusions with the Harmony software revealed that C901 (Figure 4-10 A) and C904 (Figure 4-10 B) viruses reduced the percentage of neurons displaying sense RNA Foci when comparing with control conditions.

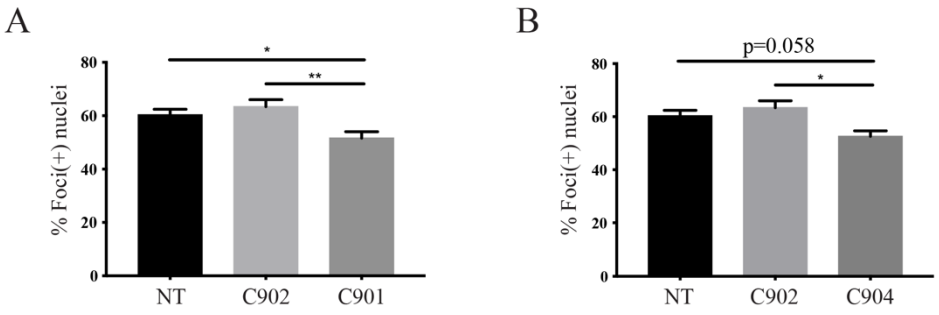


Figure 4-10. Transduction of C9-500 BAC cortical neurons with C901 or C904 reduces RNA Foci accumulation. High throughput, automated quantification using Harmony software (Perkin Elmer) shows that 6-day transduction with (A) C901 or (B) C904 therapeutic viral vectors reduces the percentage of Foci(+) nuclei by ~15% and ~19% (n=6 biological replicates), when compared with the NT (n=6 biological replicates) and C902 (n=5 biological replicates) control conditions, respectively. One-way ANOVA, *p<0.05, **p<0.01.

Considering how this decrease is also significant in relation to the control sgRNAs virus (C902), these results suggest that the editing of the C9orf72 gene is the underlying mechanism that is driving the reduction of RNA Foci. Also considering the well-studied role of sense Foci in C9orf72 pathology (Česnik et al., 2019; Cooper-Knock et al., 2014a), we can speculate that C901 and C904-mediated editing might have therapeutic benefit in C9orf72-associated diseases.

Bolstered by these results, we wondered whether this effect would have an impact in more downstream by-products such as DPRs, which result from the non-canonical RAN translation of the G4C2 repeats, and play an important role in the pathophysiology of this mutation (Goodman and Bonini, 2019). One of the biggest difficulties, however, when studying endogenous non-tagged DPRs is the high background levels and low specificity of available anti-DPR antibodies, and the inadequacy of conventional methods to study protein expression, such as Western blot. Dot blot (Cykowski et al., 2019; Hautbergue et al., 2017; Mackenzie et al., 2015) and ELISA assays (Gendron et al., 2015; Lehmer et al., 2017; Lopez-Gonzalez et al., 2016; Simone et al., 2018; Su et al., 2014) are examples of techniques that have been adapted to overcome these issues and are used widely to study the expression levels of DPRs. Considering this, we transduced cortical neurons from the C9-500 BAC model at 2 days *in vitro* (DIV 2) with our different viruses for a total of 6 days until DIV 8, after which we ran Dot blots on protein extracts and probed with an antibody against poly-(GP). Interestingly, we observed a significant reduction in poly-(GP) accumulation in neurons transduced with C901 and C904 viruses when compared with the C902 control condition (Figure 4-11). Also, consistent with the ELISA results, the levels of poly-(GP) in non-transgenic neurons were considerably lower than in transgenic, demonstrating the specificity of our assay.

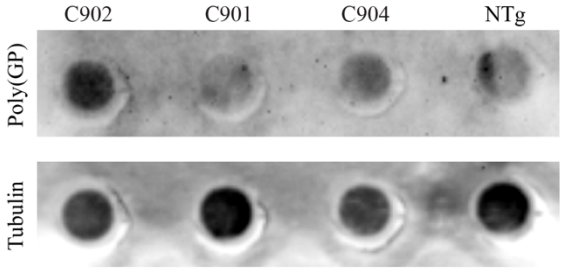


Figure 4-11. Transduction of C9-500 BAC cortical neurons with C901 and C904 reduces poly-(GP) DPRs accumulation. Dot blot analysis from protein extracts of C9-500 cortical neurons showed that 6-day transduction with C901 and C904 therapeutic viruses reduces poly(GP) DPRs expression in these cells when compared with neurons transduced with C902 control. Non-transgenic (NTg) neurons display low signal for poly-(GP).

Due to technical and user variability, most authors are hesitant to quantify Dot blot results by measuring signal intensity (Cykowski et al., 2019; Hautbergue et al., 2017; Mackenzie et al., 2015), as opposed to what is done routinely with Western blots. For the same reasons, we did not quantify the poly-(GP) signal in our membrane, but we did additional dot blots to confirm that the C901 and C904-mediated decrease was reproducible. Indeed, two more dot blots (Figure 4-12 and **appendix 5**) suggested that treatment with C901 and C904 might decrease accumulation of poly-(GP) in primary neurons when compared with control conditions. We have also attempted to study the impact of our therapeutic strategies on other DPR species such as poly-(GA) and poly-(GR) but have failed to see a specific signal by Dot blot and ELISA with existing antibodies.

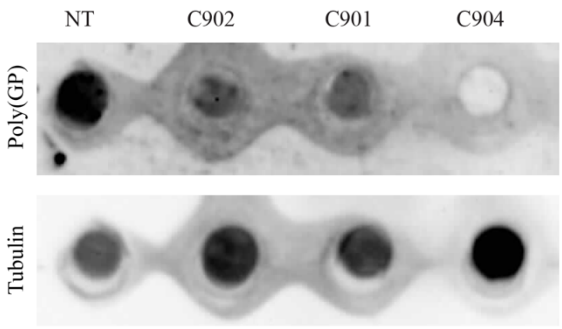


Figure 4-12. Complementary Dot blots suggest decrease of poly-(GP) upon treatment with C901 and C904 is reproducible. Dot blot analysis from protein extracts of C9-500 cortical neurons showed clearly that 6-day transduction with C904 therapeutic virus reduces poly(GP) DPRs

expression in these cells when compared with control conditions. Transduction with C901 reduced poly-(GP) compared to the non-treated condition although its comparison with C902-treated samples is not clear. A leak in the well used for C902 might explain its decrease in signal. An additional dot blot was done that replicates these findings and is shown in **appendix 5**.

4.4 Discussion

In this chapter, we report how we took our project one step further by testing our experimental therapies in cells with a disease-associated C9orf72 expansion for the first time. Indeed, we showed how the C9-500 BAC mouse is the ideal model to study gene modulation of C9orf72 due to its genetic design, and how primary cortical neurons from this model constitute a powerful *in vitro* system for initial testing. We then demonstrated that AAV9-delivery of C901 or C904 is capable of editing an expanded C9orf72 allele derived from a patient, leading to the reduction of disease pathological hallmarks like RNA Foci and poly-(GP) DPRs as a direct result of this editing.

One of the first steps we took was to sequence the flanking regions of the G4C2 expansion in this model to confirm for complementarity of the designed sgRNAs to their target sequences in the C9orf72 (Figures 4-1 and 4-2). Indeed, our concern stems from the high genetic variability between patients, particularly in intronic regions (Altshuler et al., 2010; Auton et al., 2015), that might hinder the binding of sgRNAs to this region. Even though SaCas9 can tolerate mismatches of its sgRNAs to their targets (Ran et al., 2015), these would significantly decrease the editing efficiency and limit the therapeutic benefit. However, even though this could pose an obstacle to clinical application, we believe it could be eventually overcome. As we demonstrated in this chapter, the PCRs we designed to amplify the regions flanking the G4C2 region can work even in the presence of the repeat expansion, which would allow us to quickly screen candidate patients' genomes before treatment administration. In addition to this, one can speculate that our constructs are designed in such a way that would allow us to adapt to the unlikely scenario of a patient with a genetic mutation in the targeted C9orf72 sequence. In this situation, the sgRNA(s) sequence(s) could be altered to accommodate the patient-specific SNP or indel in question, and the new sgRNA(s) could then be quickly cloned in the

main vector using the sgRNA cloning cassettes, followed by AAV9 production. Also, in the event that our experimental therapies would reach clinical application, a project could be designed to compile the most common genetic mutations in the sgRNAs targeted sequences in the C9orf72-carrier population. This would allow us to prepare in advance with a pool of alternative sgRNAs that could be cloned and ready for clinical-grade AAV9 production, following additional pre-clinical safety studies.

Another preliminary step we took was to investigate whether and how well our viruses were able to transduce and induce transgene expression in cortical neurons. Our primary concern related to the use of the EFS core promoter, despite some good evidence to support its use. Indeed, EFS1a-driven plasmid expression from transfections into HEK293T cells worked well for us (chapter 3), and the EFS1a promoter has been successfully used with lentiviruses and AAVs to express transgenes in a variety of tissue and cells (Heckl et al., 2014; Hoffmann et al., 2017; Kim et al., 1990; Nelson et al., 2010; Tabebordbar et al., 2016; Vora et al.). However, its conjugation with AAV in neuronal populations has produced mixed results, as described in the literature (Holehonnur et al., 2015; Kumar et al., 2018; Nishiyama et al., 2017). Nevertheless, we observed that AAV9 delivery of an EFS1a-GFP transgene was able to transduce cortical neurons from the C9-500 BAC model with approximately 60% efficiency, showing that the promoter was adequate for these experiments (Figure 4-3). A different concern we had was the larger size of the CRISPR transgenes used, and whether this could impact the viruses' transduction rate. The results we described above (Figure 4-5) suggest that indeed C901, C902 and C904 transduced C9-500 BAC cortical neurons less efficiently than the AAV9-EFS1a-GFP, which could be due to the larger transgene size. However, calculating transduction efficiency for the CRISPR viruses relies on doing immunocytochemistry, instead of direct GFP expression and imaging, which might also explain the disparity of results when comparing with AAV9-EFS1a-GFP.

Having established the C9-500 BAC cortical neurons as a robust *in vitro* system for this project, we proceeded to investigate the therapeutic potential of our approaches, despite prospective complications that we had anticipated. In fact, it was unclear whether editing a C9orf72 gene with a full G4C2 expansion would be possible, due to its highly unstable nature and complex secondary structures. Surprisingly, though, we were encouraged by the C901 treatment's ability to excise this expansion, promoting the re-ligation of the gene around the predicted cut sites (Figure 4-8). We suspect that the design of the sgRNAs to bind ~90-100 bp outside of the repeat region was important for this success. Although the calculated efficiency for this

correction was ~1.5%, this is likely an underestimation. Indeed, the qPCR assay used was designed based on the most common cut and re-ligation event we observed during this project and cannot detect alternative editing conformations. A future experiment that is being planned consists in designing a more permissive assay, by using a qPCR probe that can still bind to the target DNA even with some mismatches. This should allow us to more accurately estimate C901-mediated editing. Perhaps even more encouraging, C904 treatment was also able to edit the C9orf72 gene, showing that SaCas9 could still access its target site despite the close proximity of one of the sgRNAs to the expansion (Figure 4-9).

In addition to editing a mutated C9orf72 allele, C901 and C904 gene transfer reduced the accumulation of sense RNA Foci in treated cortical neurons. Importantly, this reduction was significant when comparing with C902, linking this effect to the therapeutic sgRNAs designed. Surprisingly, C904 treatment also reduced RNA Foci in these cells, despite the presence of the repeat expansion. However, this strategy actually excises an entire ~90 bp sequence directly upstream of the G4C2 repeat region, and not just the critical 'CUG' start codon. Therefore, one can speculate that this targeted sequence might contain other regulatory elements important for the transcription of the expansion, which might explain the reduction in RNA Foci. Also, it is worth noting that the RNA Foci were imaged using an Opera phenix high throughput confocal microscope (Perkin Elmer) and quantified automatically using the Harmony software (Perkin Elmer), constituting a robust and unbiased screening method that can be applied to other therapies being tested for C9orf72-linked ALS/FTD.

The last set of experiments described in this chapter focused on the effect of our therapeutic strategies on the accumulation of DPRs. This investigation was sometimes frustrated by the inadequacy of conventional protein quantification methods, such as Western blot, due to the multiple sizes and fragmentation of endogenous DPRs creating smears that cannot be properly quantified (data not shown). In some cases, these can still be used for qualitative analysis (Mackenzie et al., 2015) but would not be appropriate to compare differential expression across treatments. Also, in our hands, antibodies against endogenous DPRs have often proven to produce inconsistent and non-reproducible data or give high background signal (data not shown). However, since it does not rely on protein separation by size, Dot blots have worked much better for us in the past to compare expression of exogenous DPRs, and even endogenous poly-(GP), across different treatment groups (Hautbergue et al., 2017). Therefore, we used Dot blots to show that the accumulation of poly-(GP) in C9-500 BAC transgenic cortical neurons is reduced in C901 and C904-treated cells, respectively (Figures 4-11 and 4-12, and **appendix**

5). Even though caution is warranted when analysing and interpreting Dot blots, an important observation is that C901 and C904-mediated reduction in poly-(GP) looks drastic, when comparing with the slight decrease they caused in RNA Foci and with the seemingly low rate of precision editing induced. To investigate this issue further, we are planning to study this reduction by ELISA, which is a more sensitive technique. However, we can speculate that C901 and C904 are mutagenizing the sequences flanking the expansion, even when this editing does not result in a perfect deletion and re-ligation of the gene. In other words, we suspect that the editing rate induced by our strategies might be considerably higher that what we refer to as precision editing, and that these disruptions are interfering with important RAN translation elements. We hope that, in the future, better antibodies and more sensitive assays will also allow us to study the effect of our therapeutic strategies on different species of endogenous DPRs, some of which are considerably toxic, such as arginine-containing DPRs (Kwon et al., 2014; Mizielinska et al., 2014; Moens et al., 2019; Wen et al., 2014; Zhang et al., 2018). Despite being largely non-toxic, however, the observed reduction in poly-(GP) is already very encouraging to us, considering the important characteristics of this DPR species as a pharmacodynamic marker for C9orf72-ALS/FTD (Simone et al., 2018). Indeed, studies in humans detected stable levels of poly-(GP) in the CSF of C9orf72 expansion carriers, while decreasing G4C2 RNA with ASOs *in vitro* and *in vivo* induced a paralleled reduction in these DPRs, confirming poly-(GP) as an attractive disease biomarker (Gendron et al., 2017; Lehmer et al., 2017). As for detecting other DPR species, we have so far done a preliminary analysis by Dot blot and ELISA on C9-500 BAC cortical neurons and failed to see a positive poly-(GR) signal in transgenic neurons (ELISA quantification shown in **appendix 6**).

To summarise, this chapter describes the characterisation of C9-500 BAC cortical neurons as a robust model to study C9orf72 gene modulation. We subsequently took advantage of this model to show how C901 and C904 AAV9 viruses are capable of editing an expanded human C9orf72 allele in these primary neurons in a reproducible manner. We have also shown how this editing impacted direct by-products of the G4C2 expansion, namely by reducing RNA Foci and poly-(GP) DPR accumulation *in vitro*, which also constitute proposed mechanisms of C9orf72 pathophysiology. These results open a new therapeutic avenue that we were keen to explore further, which includes investigating their efficacy *in vivo*.

5. *In vivo* efficacy study

5.1 Introduction

In chapter 4, we described how our therapeutic constructs can be expressed from the clinically successful AAV9 virus and edit the human C9orf72 gene in primary neuronal cells, decreasing key by-products of the G4C2 expansion. Considering the long-term goal of clinical application, it was now paramount to investigate whether these constructs could transduce key cell populations and edit the C9orf72 gene *in vivo*. In parallel, we also aimed to assess whether C901 or C904 had the potential to ameliorate C9orf72-ALS/FTD symptoms that could justify clinical translation. Although the viruses had shown promise *in vitro*, we could anticipate at least one major difficulty *in vivo*, relating to the viral dose used. Indeed, while we used a considerably high MOI in cultured neurons, described in the previous chapter, the viral load per cell will be inevitably lower in a living organism, which could significantly dilute the effect of the editing strategies. However, we hoped that this issue would be counteracted by a significantly longer transduction time *in vivo* than what could ever be accomplished in cultured cells, giving more time for the transgene to express and the editing to occur.

As for the model of choice, the C9-500 BAC mouse created by Liu and colleagues (Liu et al., 2016) is the only one so far to have reported major C9orf72-ALS/FTD features including paralysis, neurodegeneration, decreased survival, together with expression of RNA Foci and DPRs. Furthermore, this model has recently been used successfully by the same team to validate an antibody-based as well as a small molecule-based therapy for C9orf72-ALS/FTD (Nguyen et al., 2020; Zu et al., 2020). For these reasons, we decided to use the C9-500 BAC mice for our *in vivo* proof-of-concept.

Therefore, in this chapter, we describe the multiple tests and experiments done on this mouse model to determine the therapeutic potential of our proposed editing strategies. Considering its potentially long duration, with mice developing symptoms until 40 weeks of age or older (Liu et al., 2016), we started to recruit animals for the behavioural study first. Unfortunately, in our hands, these assessments were frustrated by a severe epileptic syndrome associated with the background strain of the model (FVB/N), which confounded any eventual C9orf72-associated symptoms. We then report how we took advantage of the strong genetic and molecular features of this model to investigate the biodistribution of our viruses, their ability to edit the C9orf72 *in vivo* and its impact on by-products of the expansion.

5.2 Behavioural testing of C9-500 BAC mice is hampered by a background-associated seizures syndrome

According to the previous characterization of the C9-500 BAC model (Liu et al., 2016), female hemizygous are the most susceptible to severe symptoms, for which reason we decided to use only females in our *in vivo* experiments. We started this study by injecting transgenic mice through the tail vein with a high viral dose of either C902 (control virus), C901 or C904 at 3-4 weeks of age, and kept a group of transgenic non-treated (NT) as well as a group of non-transgenic (NTg) animals. The choice of the route of administration and timing of injections was based on the following reasoning. Firstly, treating mice at neonate stage for an ageing-related disease would not be applicable to humans, for which reason injecting animals at adult stage and closer to the onset of symptoms is more relevant for clinical translation. Also, although intrathecal injections of genetic material in humans are a reliable route of administration (Finkel et al., 2017; Miller et al., 2020, 2013; Mueller et al., 2020), equivalent procedures in adult mice are highly invasive and technically challenging due to the small size of the animals. In fact, in adult mice, tail vein injections are technically much less demanding, less invasive and therefore safer to the animal. Finally, according to the literature, AAV9 can cross the blood-brain-barrier and transduce CNS tissue even when administered intravenously, both in humans (Mendell et al., 2017) and mice (Foust et al., 2009).

This study had a total duration of 63 weeks, with the mice being subjected to previously validated behavioural tests including the open field and grip strength tests. In addition to these, other biological parameters were also carefully analysed, such as body weight, key electrophysiological hallmarks, or observation of clasping, seizures and other possible symptoms. With the combination of these tasks, we envisaged to fulfil two main objectives throughout the study. First, we aimed to do an in-house characterisation of this mouse model and try to determine whether mice with a mild or more severe phenotype (Liu et al., 2016) could be distinguished before the onset of symptoms. Our second objective was to assess whether our therapeutic strategies would ameliorate any of the symptomatology of these mice.

The earliest noteworthy observation we made in the study was a striking hyperactive behaviour exhibited by some of the animals in their cages, which would sometimes even start before 10 weeks of age. Wondering if this behaviour could be a useful early indicator of milder or more severe phenotype, we attempted to quantify it using a variation of the open field test (Herranz-Martin et al., 2017). This confirmed that indeed a subset of the animals showed striking hyperactivity, consisting of repetitive running in circles, always to the same side, resulting in a considerably higher number of crossings than their littermates (Figure 5-1, **supplementary videos S1 and S2**). However, this trait was not specific to transgenic animals, and there was no statistical difference between any of the analysed groups at 10-12 weeks (Figure 5-1 A), 18-20 weeks (Figure 5-1 B) or even at a later timepoint (available in **appendix 7**).

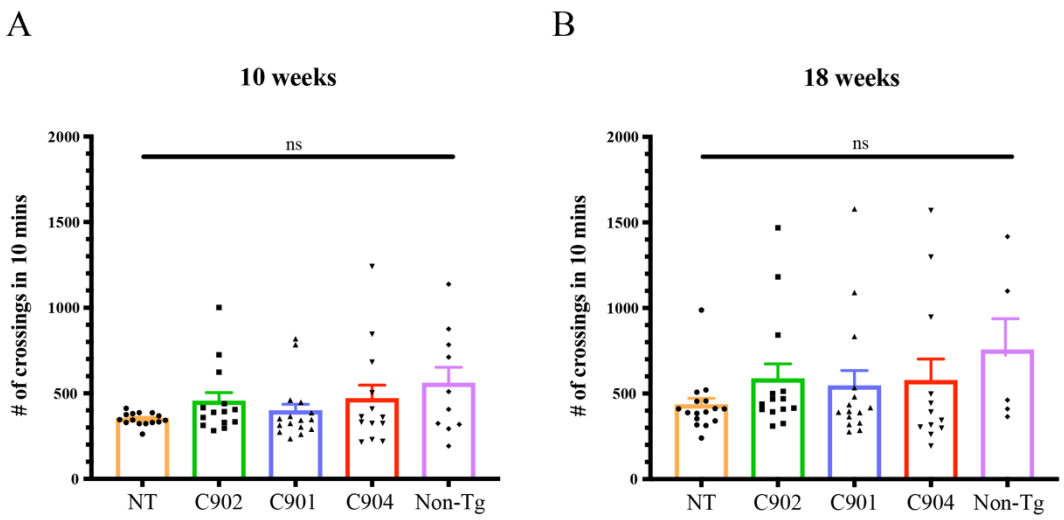


Figure 5-1. A subset of transgenic and non-transgenic C9-500 BAC mice display a hyperactive behaviour and erratic circular motion that is not transgene-specific. Open-field test at (A) 10-12 weeks and (B) 18-20 weeks reveals that a subset of C9-500 BAC mice develop early signs of a marked hyperactive behaviour, characterised by circular running movements, always to the same side (**Supplementary videos S1 and S2**). This behaviour is not specific to transgenic animals and does not change with viral treatments. NT = non-treated. (A) NT = 15; C902 = 14; C901 = 16; C904 = 13; Non-Tg = 10; (B) NT = 16; C902 = 14; C901 = 15; C904 = 12; Non-Tg = 5; One-way ANOVA, ns = not significant.

We also observed that, starting at 18 weeks of age until 50 weeks, some mice developed a severe neurological syndrome characterised by random combinations of symptoms that included lethargy, clasping, kyphosis, extreme weakness, accelerated breathing, demented behaviour, seizures and sharp weight loss (Figures 5-2 and 5-3, **supplementary videos S3 and S4**). In some cases, this symptomatology resulted in the death of the animal. Some other mice have in turn developed what seems to be a mild version of this syndrome, showing isolated symptoms such as clasping, seizures or social withdrawal with lack of interest in their surroundings and hyperactivity when provoked (Figure 5-3, **supplementary video S5**). Finally, a subset of these animals remained asymptomatic throughout the duration of the study (Figure 5-3).

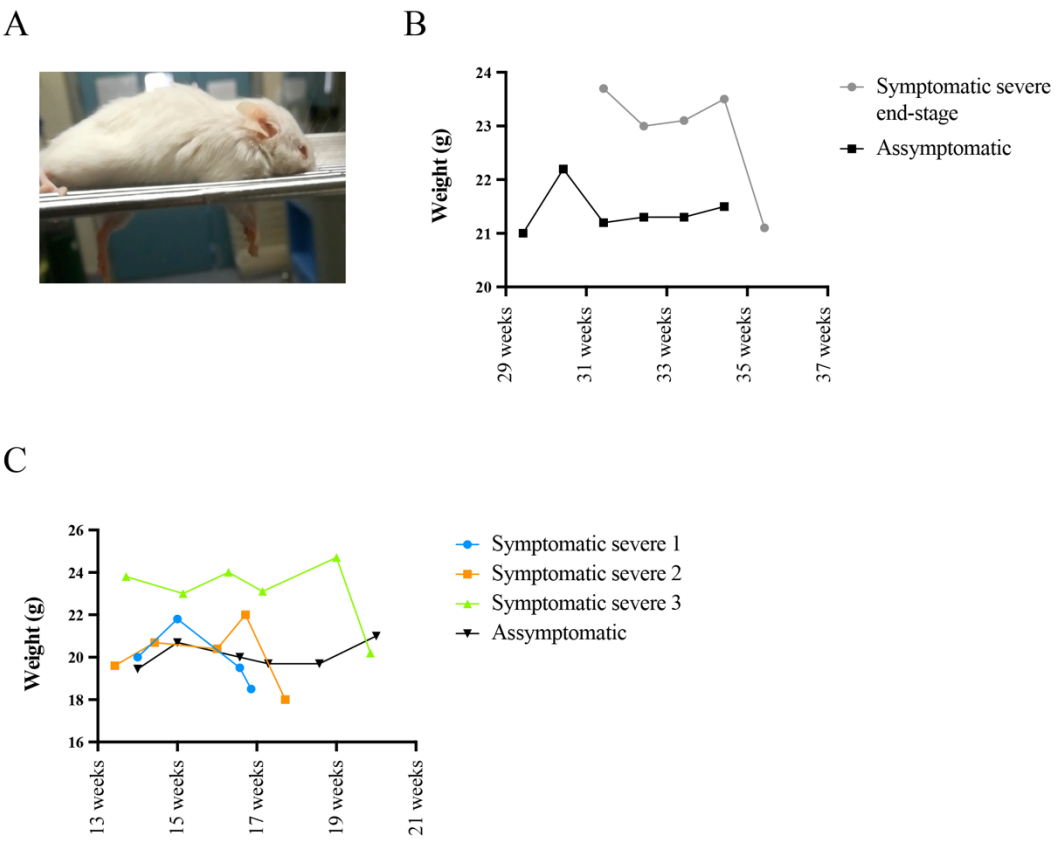


Figure 5-2. C9-500 BAC mice can develop a severe neurological syndrome with acute weight loss which can result in their premature death. (A) Representative picture of a C9-500 BAC transgenic mouse with an acute neurological syndrome characterised by lethargy, clasping, kyphosis, extreme weakness, accelerated breathing and sharp weight loss which can result or not in premature death. (B) Weight graph of a severely symptomatic animal with acute weight loss that resulted in its early death. The weight of an asymptomatic mouse of the same age is given for comparison. (C) Graph plotting the weight of 3 severely symptomatic animals suffering acute weight loss at the start of their symptoms but that recovered after this recorded drop. The weight of an asymptomatic animal of the same age is given for comparison.

The symptomatology observed by us was similar to that reported by Liu and colleagues (Liu et al., 2016), as well as Nguyen and collaborators (Nguyen et al., 2020), from the same team that initially characterized the model. However, in our hands, non-Tg littermates have developed the same symptoms (**Supplementary videos S6 and S7**) with comparable frequency as transgenics, treated or non-treated, and no statistical differences were observed between the groups, except for a tendency for less severe cases in C901- and C904-treated mice (Figure 5-3). Taken together, this data seems to suggest that the observed phenotype is not caused by the expanded C9orf72.

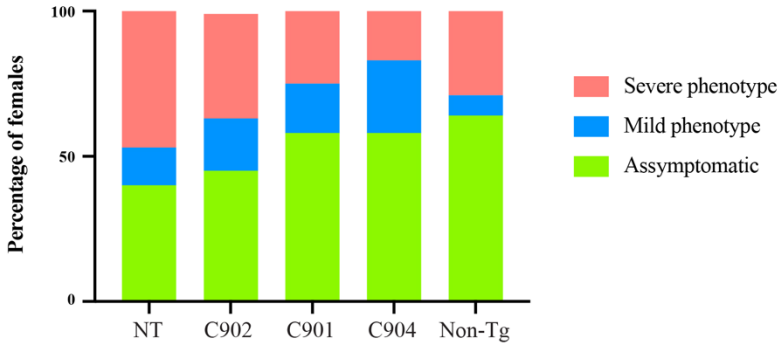


Figure 5-3. The neurological symptoms observed in this model is not exclusive to transgenic animals and does not change significantly with virus treatment. Percentages of mice with severe symptoms, mild symptoms or asymptomatic in each different treated group throughout the entire duration of the study (63 weeks). Mice that died inexplicably without obvious symptoms before 40 weeks of age were excluded from this analysis. There was no statistical difference between the different conditions. NT = non-treated. NT = 15; C902 = 11; C901 = 12; C904 = 12; Non-Tg = 14; Chi-square test, ns = not significant.

Considering the previous reporting of reduced grip strength in this model (Liu et al., 2016), and the extreme weakness observed in some of our symptomatic mice, we decided to attempt quantifying this trait. To assess limb strength, we started by choosing three timepoints corresponding to a compromise between the highest number of animals with symptoms, and a period in the study with enough surviving mice. We then employed a variation of the grip strength test at these timepoints, using an inverted grid (Deacon, 2013; Gleitz et al., 2017) which, as predicted, some mice were unable to perform. However, the test did not reveal any significantly statistical changes between any of the groups analysed (Figure 5-4).

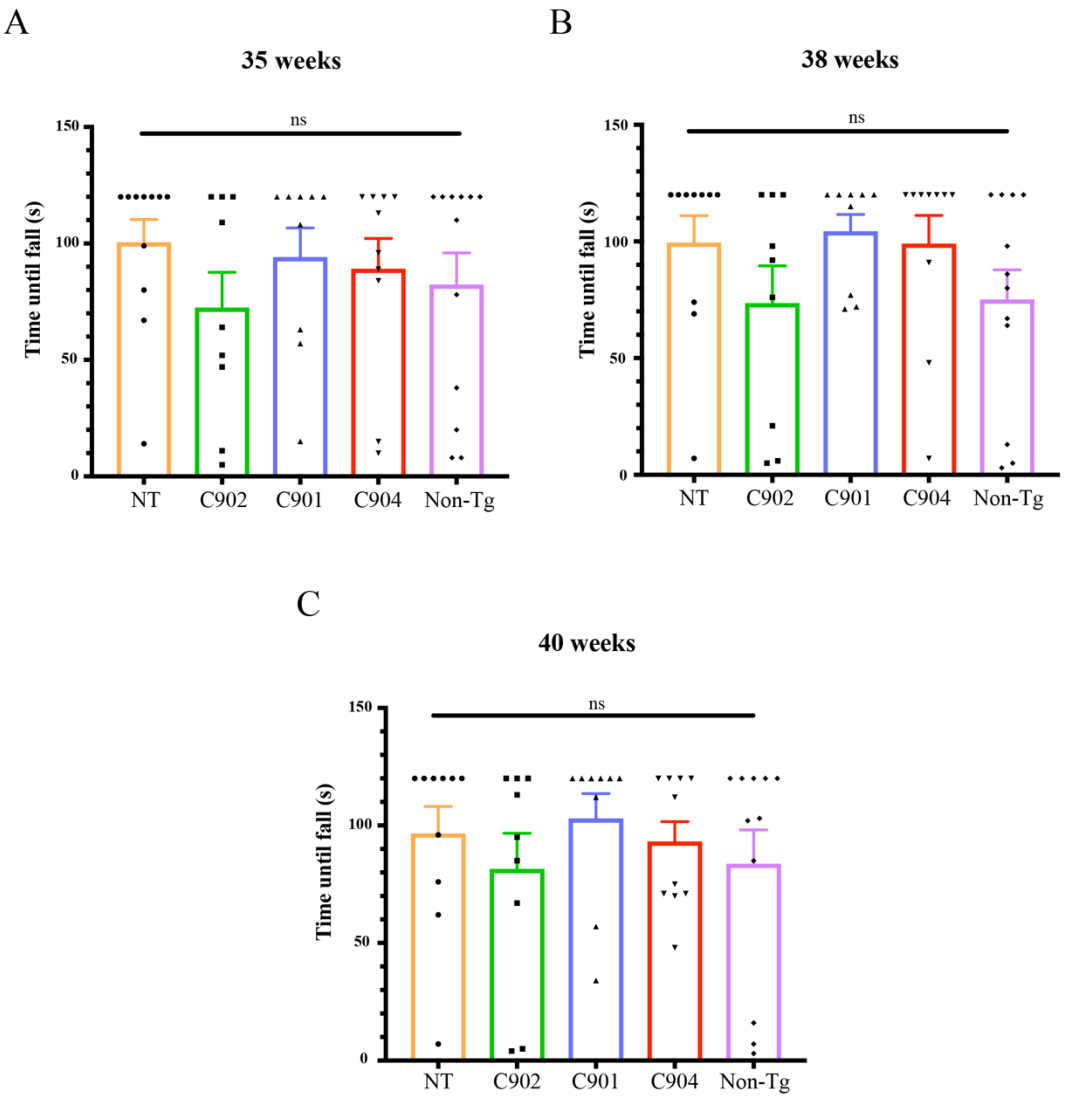


Figure 5-4. Transgenic and non-transgenic C9-500 BAC mice do not show differences in limb strength/weakness across treated groups. A variation of the grip strength test with an inverted grid at (A) 35-37 weeks, (B) 38-40 weeks and (C) 40-42 weeks did not reveal significant differences in limb strength/weakness between transgenic and non-transgenic animals, or virus-treated groups. NT = non-treated. (A) NT = 11; C902 = 9; C901 = 9; C904 = 10; Non-Tg = 12; (B) NT = 10; C902 = 9; C901 = 9; C904 = 10; Non-Tg = 12; (C) NT = 10; C902 = 9; C901 = 9; C904 = 10; Non-Tg = 11; One-way ANOVA, ns = not significant.

Importantly, in our hands, transgenic C9-500 BAC mice did not develop muscle paralysis at any point in the study, which is a major clinical characteristic of classic ALS and ALS/FTD (van Es et al., 2017; Kiernan et al., 2011) and contrasts with what was initially observed (Liu et al., 2016). To investigate further, we analysed compound muscle action potential (CMAP) recordings in the hindlegs of the animals, as an indicator of motor neuron functionality in those muscles. Indeed, although this model has been reported to display TDP-43 pathology and denervation of the neuromuscular junction (NMJ) (Liu et al., 2016), they have not been analysed for electrophysiological hallmarks of ALS and ALS/FTD. Considering this, we took *in vivo* recordings from these mice, only to verify that there were no significant differences in CMAP amplitude between any of the groups at 10-12 weeks (Figure 5-5 A) or at 18-20 weeks (Figure 5-5 B). Importantly, recordings from the hindleg muscle of extremely weak symptomatic mice with a severe neurological syndrome similar to what was previously described (Liu et al., 2016), still showed no noteworthy decrease in CMAP amplitude. Recordings took at a later stage also did not show any changes (available in **appendix 8**). These results led us to suspect that there was no transgene-associated neuromuscular dysfunction in these mice.

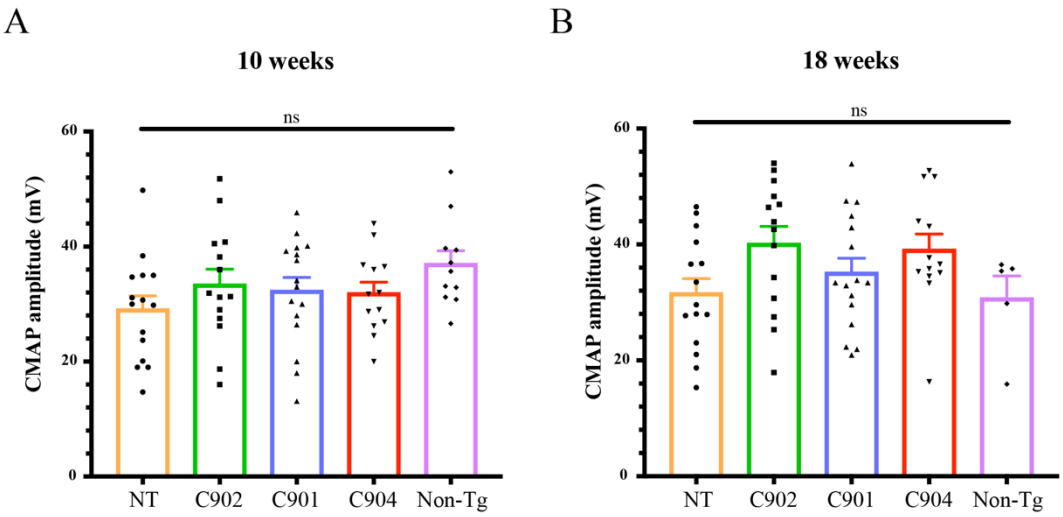


Figure 5-5. Transgenic and non-transgenic C9-500 BAC mice across treatment groups have comparable compound muscle action potential. Electrophysiological recordings *in vivo* at (A) 10-12 weeks and (B) 18-20 weeks do not show differences in CMAP amplitude between transgenic and non-transgenic animals, or virus-treated groups. NT = non-treated. (A) NT = 15; C902 = 14; C901 = 16; C904 = 13; Non-Tg = 11; (B) NT = 15; C902 = 14; C901 = 16; C904 = 13; Non-Tg = 5; One-way ANOVA, ns = not significant.

Concerning survival, one of the treated groups, transgenic mice injected with C904 virus, showed a slight tendency for survival improvement. Interestingly, however, the non-Tg littermates showed a slightly, non-significant worse survival in our hands to that of all other transgenic groups, treated or not treated (Figure 5-6), which is at odds with what was reported by the original team that characterised this model (Liu et al., 2016; Nguyen et al., 2020). This data suggests that the presence of the C9orf72 expansion is not actually decreasing the survival of these animals, and that they are more likely dying due to traits linked to the background strain. Moreover, the deaths annotated during this study were not only linked to severely symptomatic animals but were in many cases the result of sudden deaths, not specific to any particular treated group. Indeed, several non-transgenic and transgenic mice were frequently found dead in their cages without any apparent cause of death, even after autopsy, a phenomenon that had been reported previously in the FVB/N strain (Mahler et al., 1996; Rosenbaum et al., 2007).

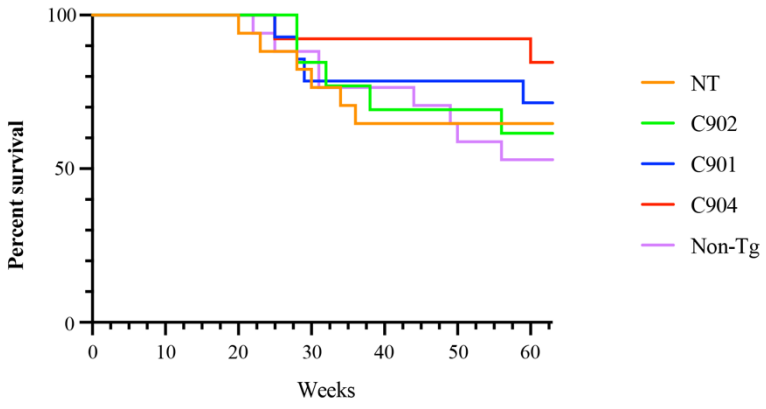


Figure 5-6. Transgenic and non-transgenic C9-500 BAC mice do not show differences in survival across treated groups. Kaplan-Meier curves for a 63 weeks survival study show that transgenic NT and non-Tg groups have comparable survival, with non-Tg mice showing a slight, non-significant tendency for worse survival. The group treated with C904 shows a tendentious improvement in survival when comparing with transgenic NT and C902-treated group, but it is not statistically significant. NT = non-treated. NT = 17; C902 = 13; C901 = 14; C904 = 13; Non-Tg = 17; Mantel-Cox (log-rank) test, Gehan-Breslow-Wilcoxon test, ns = not significant.

An important observation made during this project that might explain the origins of the background-associated syndrome was the occurrence of several episodes of epileptic seizures in these mice, briefly mentioned above. These episodes happened in 7%-25% of every studied group, suggesting they are background-associated, and not transgene-specific. This quantification needs to be cautiously interpreted and it is likely to be an underestimation, since it does not account for seizures that might have happened when the mice were not being handled. These episodes varied in intensity, but all of them included facial grimace, tonic-clonic convulsions and loss of bladder control, with the animals appearing extremely weak afterwards and unable to perform the grip strength test. Interestingly, this epileptic syndrome has been reported previously in the FVB/N background (Goelz et al., 1998; Silva-Fernandes et al., 2010; 2007; Ward et al., 2000), with some authors labelling it 'space cadet' syndrome and reporting remarkably similar symptoms to what we observed in this study. This included observations of strong tonic-clonic seizures particularly prevalent in females, resulting either in early death or subsequent social withdrawal/dementia features that gave the name to this syndrome. A closer look at our data actually revealed a remarkably strong correlation between observed seizures and the symptoms described above, with 87.5% of our mice with observed seizures also developing the additional symptoms associated with the severe phenotype. In addition to this, an internal investigation at the Jackson Laboratory also reported no phenotypic differences between transgenic and non-transgenic C9-500 BAC mice and concluded that the symptoms observed correlated with seizure episodes. Importantly, though, this investigation has not been peer-reviewed and needs to be interpreted with caution. However, considering all of our results put together and what was previously reported in the literature, we were forced to conclude that, in our hands, the neurological symptoms observed in these mice are not specific to the transgene and are likely masked by, or the result of, a severe background-

associated epileptic syndrome. Therefore, the behavioural parameters observed in these mice are inappropriate to study the efficacy of our therapeutic strategies.

5.3 Intravenous administration of C901 and C904 can transduce CNS and peripheral tissue and edit the human C9orf72 *in vivo*

Studying the behavioural impact of C901 and C904 in the C9-500 BAC model was not possible, as described above. However, it was still warranted to investigate whether our therapeutic viruses could reach and transduce disease-relevant tissue in these mice and edit the human C9orf72 *in vivo*. With this objective in mind, we started by studying the viral biodistribution of our AAV9 viruses in these mice by injecting a small number of C9-500 BAC females in the tail vein and subsequently analysing their tissue for the presence of the transgene 4-6 weeks later. Importantly, qPCR detected the presence of SaCas9 cDNA in the frontal lobe, parietal-temporal lobes and spinal cord of injected mice, showing that C902, C901 and C904 AAV9 viruses were able to cross the blood-brain-barrier and transduce distinct regions of the CNS (Figure 5-7 A). Low transgene levels detected in the cerebellum suggests poor transduction of this region and a reduced chance of detecting editing events. Similar to what was described in the literature (Inagaki et al., 2006; Pacak et al., 2006; Zincarelli et al., 2008), our recombinant AAV9 viruses also showed considerably higher tropism for peripheral tissue upon intravenous administration, including the liver, the heart and muscle (Figure 5-7 B).

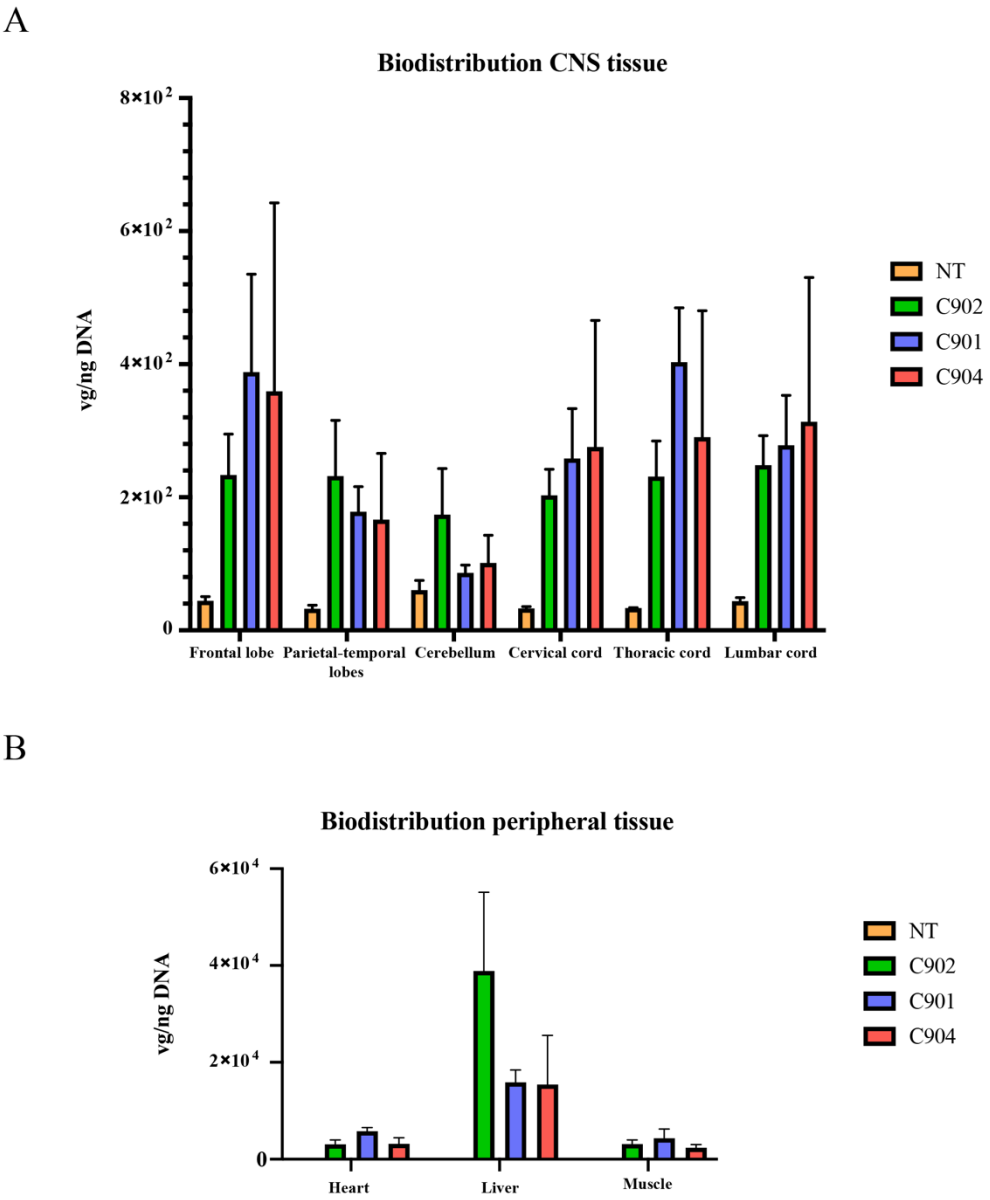


Figure 5-7. AAV9-mediated delivery of C901, C902 and C904 through the tail vein can transduce CNS, but has preferential tropism for peripheral tissue in C9-500 BAC mice. (A) qPCR on DNA extracted from tissue of 10-12 week old animals shows that the therapeutic and control viruses used can cross the Blood-Brain-Barrier and transduce different regions of the brain and spinal cord upon intravenous injection (n=4 mice per group). (B) The same qPCR with DNA extracted from different organs shows that the viruses used have preferential tropism for the liver, heart and muscle, with a significantly higher number of viral genome copies detected when compared with CNS tissue (n=4 mice per group).

Having established widespread transduction of CNS and peripheral tissue by our recombinant viruses, we set out to investigate whether this transduction resulted in editing of the C9orf72 *in vivo*. We started by analysing the DNA extracted from C901-injected mice by endpoint PCR with C9_OUT_2 primers, across the repeat region. Interestingly, this PCR detected C901-mediated excision of the G4C2 repeats in different regions of the brain, which was absent in not treated (NT) and C902-treated mice (Figure 5-8 A). As expected, there was seemingly less edited DNA in the cerebellum, possibly due to its lower transduction rate, and, surprisingly, no edited DNA in the spinal cord. As for peripheral tissue, the same PCR revealed edited C9orf72 DNA in the heart and liver of treated mice, and extremely faint amplification in the hindleg muscle (Figure 5-8 B). An equivalent editing study done on a different C901-injected mouse is also shown (available in **appendix 9**), demonstrating this editing is reproducible.

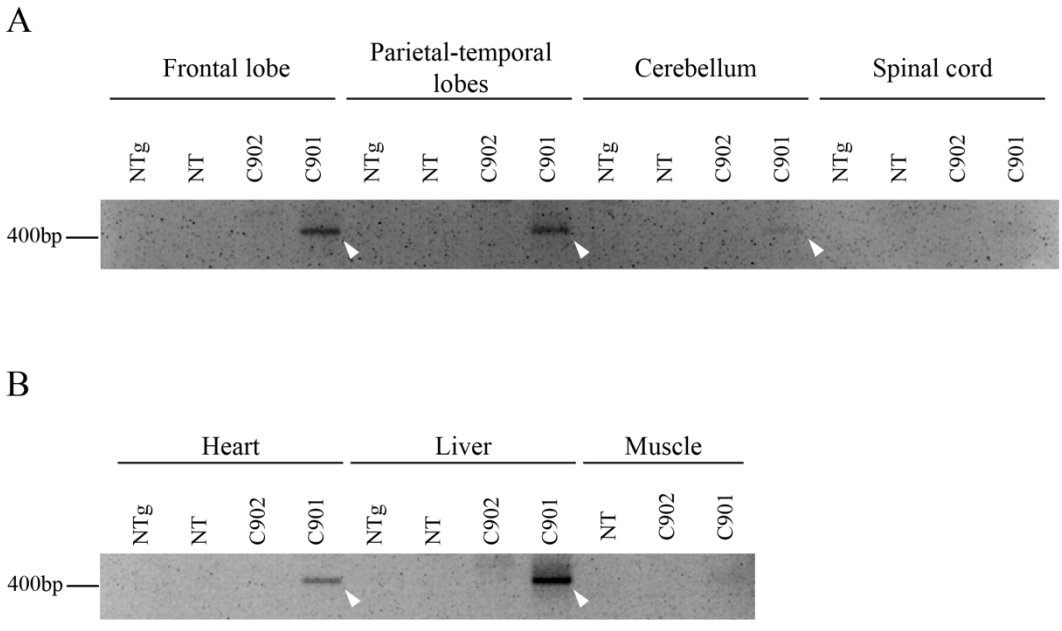


Figure 5-8. Intravenous administration of C901 can edit the human C9orf72 gene in CNS and peripheral tissue *in vivo*. (A) Agarose gel showing evidence of *in vivo* editing by C901 in the brain, but not the spinal cord, of a C9-500 BAC mouse 6 weeks after viral vector delivery in the tail vein. PCR amplification with C9_OUT_2 primers flanking the GC-rich repeats yields a product band of ~400 bp (white arrowheads) in the brain of a mouse injected with C901, which is absent in a transgenic NT and a C902-treated mouse. Non-Tg animals lack a human C9orf72 gene, for which reason the amplification with C9_OUT_2 primers does not yield a product. (B) The same PCR amplifies edited C9orf72 also in the heart and liver of the same animal (white arrowheads). Another agarose gel is shown (**Appendix 9**), with DNA from a different C901-injected mouse, showing that this editing is reproducible.

Having proved that C901 virus can edit the human C9orf72 *in vivo*, we then focused our efforts in detecting C904-mediated editing. However, to our surprise, we were able to detect C904-induced deletions only in the liver of treated mice (Figure 5-9), failing to detect edited C9orf72 in any other tissue analysed (data not shown). Although the reason for this is unclear, we can speculate that the PCR used to detect this editing might be less sensitive than the assay we use for C901-mediated editing, which exclusively amplifies edited product.

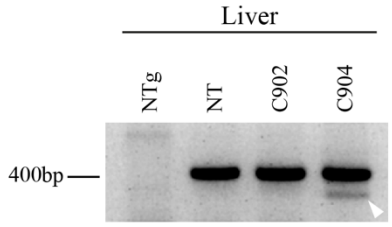


Figure 5-9. C904 can edit the human C9orf72 gene in the liver of C9-500 BAC mice upon intravenous administration. Agarose gel showing C904-mediated *in vivo* editing of the C9orf72 gene in liver DNA from a C9-500 BAC mouse. PCR amplification with C904_INT_14 primers amplifies an additional edited band in DNA from a mouse injected with C904, just

below the unedited product (white arrowhead). In contrast, transgenic NT or C902-injected mice only have unedited C9orf72.

Considering the data presented in this section, we have provided evidence that intravenous injection of the control C902 and the therapeutic viruses C901 and C904 can cross the cross the blood brain barrier and transduce distinct regions of the CNS, although maintaining a preference for peripheral tissue. We have also shown that C901 treatment can excise the G4C2 expansion from the human C9orf72 gene in the brain upon intravenous administration in C9-500 BAC transgenic mice. As for C904, its editing might be limited to the liver of treated mice, although this needs to be confirmed with more sensitive detection techniques.

5.4 Intravenous administration of C901, but not C904, can reduce poly-(GP) accumulation in the brains of C9-500 BAC mice

Having shown that C901 was capable of editing the C9orf72 gene *in vivo*, it was now warranted to check whether this editing had therapeutic potential, by investigating its effect on C9orf72-molecular hallmarks. Accumulation of DPRs is one of these key pathological hallmarks, as they constitute a direct by-product of the expansion resulting from RAN translation of the repeats, while its arginine-containing species have been proven to induce neurotoxicity in cellular and animals models (Kwon et al., 2014; Mizielinska et al., 2014; Moens et al., 2019; Wen et al., 2014; Zhang et al., 2018). However, as discussed in chapter 4, detection and quantification of endogenous non-tagged DPRs is technically challenging by conventional techniques. Although Dot blots have worked well for us *in vitro* (chapter 4), ELISA assays have actually become the gold standard for sensitive DPR detection and quantification, having been used to establish poly-(GP) as an attractive pharmacodynamic marker, for which reason this was our method of choice (Gendron et al., 2015, 2017; Lehmer et al., 2017; Lopez-Gonzalez et al., 2016; Simone et al., 2018; Su et al., 2014). For this assay, we used protein

samples from 12-week old transgenic mice, either not treated (NT), treated with control virus C902 or with therapeutic viruses C901 and C904. As a negative control, we used protein samples from 12-week old non-transgenic mice, with the samples being taken from an equivalent region of the brain, the parietal-temporal lobes, for all the mice analysed. MSD ELISA on these samples revealed that transgenic C9-500 BAC mouse brains have positive signal for poly-(GP) when comparing with non-transgenic littermates (Figure 5-10 A and B). Interestingly, mice treated intravenously with C901 showed decreased poly-(GP) accumulation in the brain when comparing with non-treated mice (Figure 5-10 A). Treatment with C904, however, did not significantly reduce poly-(GP) levels when comparing with control conditions (Figure 5-10 B), somewhat validating the previous results showing lack of C9orf72 gene editing in the brain.

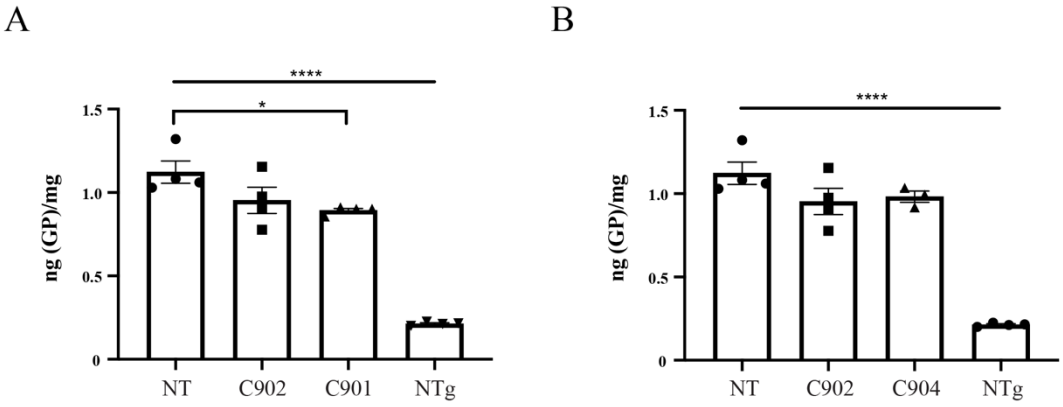


Figure 5-10. Intravenous injection of C901, but not C904, can reduce poly-(GP) DPRs in the brains of C9-500 BAC mice. MSD ELISA on brain samples from C9-500 BAC mice successfully identifies a positive poly-(GP) DPR signal in transgenic samples over non-Tg. (A) Intravenous administration of C901 can reduce the poly-(GP) accumulation in the brain when comparing with non-treated animals (n=4 mice per group). (B) Transgenic mice treated with C904 did not show a change in poly(GP) accumulation when comparing with control conditions NT = 4; C902 = 4; C904 = 3; Non-Tg (NTg) = 4. One-way ANOVA, *p<0.05, ****p<0.0001.

5.5 Discussion

In this chapter, we took the project one step forward by testing our experimental therapies in an *in vivo* mammalian system for the first time. We started by doing a behavioural study to assess whether C901 or C904 ameliorated different C9orf72-ALS/FTD symptoms, but eventually concluded that an assessment like this was frustrated by a background-specific seizures syndrome with severe neurological consequences. However, by focusing on the molecular aspects of this mouse model, we were able to draw some important conclusions regarding the biodistribution of our recombinant viruses and their ability to edit the human C9orf72 gene *in vivo* and decrease by-products of the expansion. Indeed, we were, to the best of our knowledge, the first to therapeutically excise a pathogenic G4C2 repeat region from the C9orf72 gene in the brain of an adult living mouse.

Perhaps the most surprising aspect of this chapter was the behavioural analysis of the C9-500 BAC mice. In the original characterisation of the model (Liu et al., 2016), the authors reported that ~30-35% of female transgenics develop a fast-progressing ALS-like disease, absent in non-transgenics, characterised by paralysis, decreased survival, neurodegeneration and NMJ denervation. In addition to this, the authors also described neurodegeneration in FTD-relevant areas of the brain, suggesting that these mice develop an ALS/FTD-like syndrome. This contrasts with what we observed in this study on several different aspects. In fact, although some of our mice developed a severe syndrome (**supplementary videos S3 and S4**) that is remarkably similar to what the original work also described and showed on video (Liu et al., 2016), the mice we observed were not paralysed, and in some occasions recovered their motor function entirely, instead progressing to a mild disease (**supplementary video S5**). Also, CMAP readings taken from hindleg muscle of mice at the peak of these symptoms, showing severe weakness and unable to perform the grip strength test (Figure 5-4), were still normal compared to non-transgenic littermates, suggesting no motor neuron dysfunction (Figure 5-5). Perhaps more importantly, in our hands, non-transgenic littermates could also develop these severe symptoms (Figure 5-3 and **supplementary videos S6 and S7**), suggesting that they were not linked to the C9orf72 expansion. Finally, survival data collected in our *in vivo* study revealed that non-transgenic mice had slightly worse survival to that of all transgenic groups

(Figure 5-6) which, combined with the evidence above, made us suspect that the severe symptoms observed were not a result of C9orf72-ALS/FTD. As an additional search for typical C9orf72 hallmarks, we collected hindleg muscle for NMJ staining, which has been processed and will be analysed in the near future.

The authors in the original publication (Liu et al., 2016) also described how an additional subset of transgenic females and males developed a slow-progressive, more mild disease with kyphosis, reduced activity, hyperactivity when provoked and clasping. However, in our study, despite indeed observing these symptoms, they were once more not exclusive to transgenic animals. Instead, the symptoms we described correlated remarkably with background-associated seizure episodes, a trait of FVB/N mice that had been reported before and named 'space cadet' syndrome by some authors (Goelz et al., 1998; Silva-Fernandes et al., 2010; 2007; Ward et al., 2000). These facts forced us to conclude that all the symptoms we observed throughout our *in vivo* study could be explained by this epileptic disease. As for the spontaneous death occurrences without an obvious cause, these were likely the result of devastating and lethal seizures that happened while the mice were not being handled, as was reported previously by other teams (Kohnken and Schwahn, 2016; Mahler et al., 1996; Rosenbaum et al., 2007).

Despite the contradicting evidence, the same group that initially characterised this model were able to recreate their initial findings in a second work, where they validated an antibody therapy for C9orf72-ALS/FTD (Nguyen et al., 2020). In addition to this, the same research group recently used this model to investigate the therapeutic potential of the drug metformin, studying its effects on some behavioural aspects of these mice (Zu et al., 2020). The reasons behind these discrepancies are unclear, but we were not the only ones to reach the same conclusions. Indeed, an internal investigation done by the Jackson Laboratory, although not peer-reviewed, reported very similar results to our own, with a preliminary summary of their findings published on their website (The Jackson laboratory, 2020). We also provide a direct excerpt of this summary, taken from the Jackson Laboratory website, for reader convenience (available in **appendix 10**). The road to create a reliable transgenic mouse model of expanded C9orf72 has been a difficult one from the start (Hayes and Rothstein, 2016), hindered by numerous challenges that illustrate the remarkable complexity of this mutation. A considerable part of this complexity is undoubtedly linked to the intrinsic instability of the G4C2 expansion, which results in repeat size variability between generations or even within tissues of the same individual, a phenomenon observed both in human patients (van Blitterswijk et al., 2013;

Ebbert et al., 2018; Nordin et al., 2015; Suh et al., 2015) and mouse models (Jiang et al., 2016; Liu et al., 2016). Indeed, in our own study, we have occasionally observed faint amplification of short G4C2 repeats in non-treated transgenic mice, with different sizes even within the same CNS tissue. This raises the question of how variable repeat sizes really are across distinct tissue of the same mouse, and in between generations, and what effects this might have in the associated phenotype. Another potential source of variability relates to the hypermethylation detected in the promoter (Liu et al., 2014; McMillan et al., 2015; Russ et al., 2015; Xi et al., 2013) or even the G4C2 expansion itself (Xi et al., 2015), with some authors suggesting that this can be neuroprotective. However, problems related to the instability of the repeats size, or even hypermethylation, still do not explain the disparity between different teams' results concerning the symptoms in non-transgenic littermates. A possible explanation for this might be linked to the existence of subpopulations of the FVB/N strain in different research groups worldwide, first originated by spontaneous mutations and then maintained with continuous inbreeding. Indeed, in the work done by Kohnken and Schwahn (Kohnken and Schwahn, 2016), the authors did not observe seizures in the FVB/N mice analysed by them, and suggested that the 'space cadet' syndrome has been extinct in certain cohorts around the world. Therefore, one can hypothesise that, once donated to the Jackson Laboratory, the C9-500 BAC mice from the original team (Liu et al., 2016) were backcrossed into a different subpopulation of FVB/N mice with the 'space cadet' epileptic syndrome, masking any symptoms specifically linked to the C9orf72 expansion.

Despite the issues mentioned above, this *in vivo* work with the C9-500 BAC model allowed to draw important conclusions about our proposed therapeutic strategies. In particular, the C901 approach produced exciting results since, to the best of our knowledge, we were the first to show that intravenous injection of AAV9 encapsulating SaCas9 with a pair of gRNAs can excise a G4C2 pathogenic expansion from human C9orf72 in the brain (Figure 5-8). The implications of this can be significant, as it opens the door to potentially revolutionary therapies for at least two incurable genetic conditions. However, an issue that requires careful consideration relates to the efficiency of this editing, and whether it would be therapeutically relevant. Indeed, one of our near future experiments will involve carefully studying the C901-mediated editing in the CNS and peripheral tissue using the qPCR assay described in the previous chapter and a new probe-based qPCR currently in development. Importantly, though, we did not observe C901-mediated editing in the spinal cords analysed (Figure 5-8 A), despite similar levels of transduction to those of the brain (Figure 5-7 A), suggesting this might be due

to the lack of EFS1a promoter activity in that region of the CNS. This issue decreases confidence in the direct translatability of this approach, considering the crucial role played by the spinal cord in the pathophysiology of ALS.

As for the C904 approach, it was surprising not to detect CNS editing in the animals analysed despite a similar biodistribution pattern to the other viruses (Figure 5-7 A). This could mean that either the viral vector is not introducing the desired deletion or that, more likely, there is a sensitivity issue with the detection method. Indeed, the PCR with C904_INT_14 primers has worked well for us in amplifying C904-edited C9orf72, as described in previous chapters and this one (Figure 5-9), but it might not be sensitive enough to detect weak editing activity. A possible explanation for this is that, contrary to the C9_OUT_2 PCR used for the C901 approach, the edited DNA has to compete with unedited molecules for the primers during the amplification step.

The final experiment described in this chapter focused on assessing the impact of C901 and C904 in the accumulation of DPRs *in vivo*. In the previous chapter, C901 and C904 were shown to decrease poly-(GP) in primary neurons from the C9-500 BAC mice, but it remained unknown whether this effect would be replicated *in vivo*. As discussed before, accurate and reliable quantification of endogenous DPRs is technically challenging, but we and others have shown that this can be achieved with dot blots or ELISA, especially for poly-(GP) (Gendron et al., 2015, 2017; Hautbergue et al., 2017; Lehmer et al., 2017; Lopez-Gonzalez et al., 2016; Simone et al., 2018; Su et al., 2014). Studying poly-(GP) is also of particular interest since it is translated from both the sense and antisense strands and it can be detected in the CSF of human symptomatic and asymptomatic C9orf72 expansion carriers using MSD immunoassays (Gendron et al., 2017; Lehmer et al., 2017). In addition to this, the expression levels of poly-(GP) DPRs are highly stable and do not change with time or disease progression, while decreasing G4C2 RNA induced a comparable reduction in these DPRs, making poly-(GP) a very attractive pharmacodynamic marker for C9orf72-ALS/FTD (Balendra et al., 2017; Gendron et al., 2017; Lehmer et al., 2017). For this experiment, we used an MSD ELISA assay to show how intravenous injection of C901 decreased poly-(GP) accumulation in the parietal-temporal region of the brain when compared with the non-treated animals (Figure 5-10 A). Importantly, though, this difference was not significant when compared with C902-treated mice, which showed high variability in detected DPRs levels. For this reason, it will be important to increase the number of animals treated in the near future, to determine if the poly-(GP) decrease is a direct result of C901-mediated editing. In mice treated with C904, though,

there was a tendentious decrease in poly-(GP), but this difference was not statistically significant (Figure 5-10 B), again raising doubts about the ability of this virus to edit the C9orf72 gene in the CNS. We hope that, in the future, improved antibodies will allow us to study the effect of our therapeutic approaches on other DPR species, mainly the ones that have been shown to be cytotoxic, such as arginine-containing DPRs. So far, a poly-(GR) ELISA assay has failed to detect a positive signal in brains from transgenic C9-500 BAC mice (data not shown). At this point we have also processed and cut tissue from this animal study in the cryostat to investigate DPR and RNA Foci expression by histology, but our plans have been compromised by the COVID-19 crisis and will be re-evaluated in the near future.

To summarise, this chapter describes how we tackled the third objective/aim of the project, consisting of validating our editing strategies *in vivo*. Our attempts to study the impact of these approaches on the described behavioural impairments and ALS/FTD-like symptoms of the C9-500 BAC mice were frustrated by a severe background-associated epileptic syndrome in our colony. We have also shown that C901 showed the most promising results *in vivo*, by excising the G4C2 expansion from the C9orf72 gene in the CNS and peripheral organs, and reducing poly-(GP) DPRs, a potential pharmacodynamic marker of C9orf72, in the brain upon intravenous delivery. Despite the exciting and highly innovative data presented, we were convinced that the potential therapeutic effect of our strategies could be optimised. In particular, the biodistribution study revealed a strong preferential tropism of our viruses for the liver, heart and muscle (Figure 5-7), potentially decreasing their availability for more disease-relevant tissue like the CNS. In addition to this, the activity of the EFS1a promoter used might be the underlying cause for the lack of C901-mediated editing in the spinal cord of treated mice (Figure 5-8 A). It was these important observations that informed the key changes we describe in the next chapter, in our pursuit of a more effective therapy for C9orf72-linked ALS and FTD.

6. Dual-vector PHP.eB approach

6.1 Introduction

In chapter 5, we described the testing of our proposed therapeutic strategies in the C9-500 BAC mouse model, having shown important evidence of *in vivo* editing of human C9orf72 and a decrease in accumulation of poly-(GP) DPRs. Despite the exciting developments, we were wondering whether optimisations could be made to boost the efficiency and maximise the therapeutic potential of our approaches.

The inability to detect C901-mediated editing in the spinal cord, despite similar levels of viral load to that of the brain, was of particular concern to us, much like the apparent lack of C904-mediated editing in the CNS (chapter 5.3). We theorised that the transgene expression driven by the EFS1a promoter might not be strong enough for appreciable editing in the CNS, particularly in the spinal cord. Taking this into consideration, we decided to use the ubiquitously active CMV promoter, due to its well-documented ability to mediate strong transgene expression in every tissue, including the CNS, upon AAV delivery (Gholizadeh et al., 2013; Gray et al., 2011; Lukashchuk et al., 2016; Valori et al., 2010; Zincarelli et al., 2008).

In parallel, we speculated that the delivery of the therapeutic constructs could also be optimised. Indeed, the biodistribution data of our recombinant AAV9 viruses showed significant preferential tropism for peripheral tissue (chapter 5.3), decreasing their availability to transduce broader areas of the CNS. Considering this, we decided to use a newly described AAV capsid, PHP.eB (Chan et al., 2017), which was engineered through an *in vivo* capsid selection method from PHP.B, its parent capsid (Deverman et al., 2016). In this innovative study, AAV PHP.eB was shown to be strikingly neurotropic, being able to cross the blood-brain barrier (BBB) and transduce widespread CNS areas without increased peripheral transduction, even when administered intravenously in adult mice (Chan et al., 2017). At the time of planning these experiments, there were concerns about the neurotropic properties of the parent capsid, PHP.B, being limited to particular mouse strains such as C57BL/6 or FVB/N (Hordeaux et al., 2018; Matsuzaki et al., 2019), but no information existed about the enhanced PHP.eB. Therefore, we hoped that the combination of this highly neurotropic capsid with a strong ubiquitous promoter would be a good starting point for optimising the therapeutic potential of the proposed editing strategies.

In this chapter, we describe the rationale behind the key changes introduced in our approaches, and their practical implementation. Namely, we report on using a new dual-vector paradigm inspired by research data from previous chapters of this thesis as well as published literature. We then show how these new approaches have enhanced transduction efficiency in cultured cortical neurons leading to a significant reduction of RNA Foci accumulation. Finally, we provide preliminary evidence for widespread C9orf72 *in vivo* editing in the CNS upon intravenous administration.

6.2 A dual-vector approach using AAV PHP.eB mediates stronger transduction and increased therapeutic potential *in vitro* and *in vivo*

Prior to testing our modified therapeutic vectors, expressed through the CMV promoter and delivered by the new PHP.eB capsid, there was an important aspect that warranted consideration. Indeed, the bigger size of the CMV promoter (~580 bp) in relation to the EFS1a (~215 bp) meant that the therapeutic constructs would exceed the maximum packaging size of AAV viruses (~4.7 Kb). To address this issue, we decided to try a dual-vector paradigm strategy for both our editing approaches and renaming them dual-vector C901 (dvC901) and dual-vector C904 (dvC904). In this new paradigm, dvC901 and dvC904 use the same sgRNAs and edit the C9orf72 gene in an identical way to their single-vector format, but this editing is achieved through expression of two separate constructs, each one expressing SaCas9 trough the CMV promoter and one U6-driven sgRNA. The idea to attempt this format was built on our previous research in which we showed that co-transfection of two CRISPR vectors expressing a combination of sgRNAs could edit the C9orf72 gene (chapter 3.2). It was also inspired on the previous work by Tabebordbar and colleagues (Tabebordbar et al., 2016), whose data suggested that using a dual-vector system with the strong CMV promoter was a more efficient editing tool than a single-vector approach with the EFS1a.

Having shown the suitability of C9-500 BAC cortical neurons to test our therapeutic approaches (chapter 4), we hoped that this *in vitro* system would also constitute an excellent starting point to test any optimised strategies. With uncertainty surrounding the neurotropic

properties of the PHP.eB parent capsid (Hordeaux et al., 2018), and considering how it had only been described for *in vivo* use, we first focused on checking whether these new viruses would transduce C9-500 BAC cortical neurons. We therefore produced AAV PHP.eB-CMV-GFP and calculated its titer using qPCR (Table 6.1), after which we transduced our target neurons, verifying that these AAV viruses mediated widespread and strong expression of GFP (Figure 6-1).

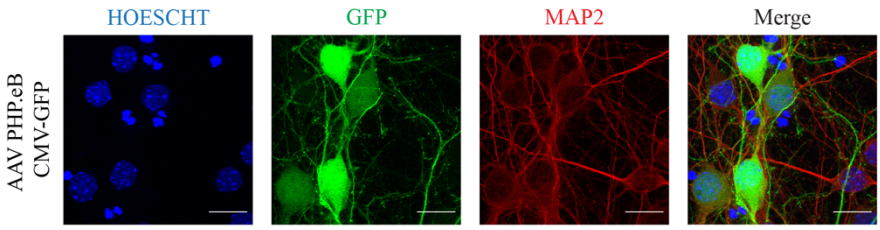


Figure 6-1. AAV PHP.eB viruses can mediate GFP expression in cortical neurons from C9-500 BAC mice. Confocal microscope images showing how AAV PHP.eB viruses can transduce cortical neurons cultured from C9-500 BAC mice and induce robust expression of GFP 6 days after transduction, demonstrating that this capsid with CMV promoter can be used to mediate transgene expression in these cells. MOI = 400'000 vg/cell. Scale bar = 20 µm.

Taking this preliminary data into account, we proceeded to make three large-scale productions of AAV PHP.eB from constructs cloned previously (chapter 3.2), all consisting of SaCas9 expressed from the CMV promoter and one of three different U6-driven sgRNAs, namely T4I, T4D or T5B. Then, according to the dual-vector paradigm explained above, the co-transduction of the CRISPR constructs expressing T4I and T5B sgRNAs was designated dvC901, and the combination of T4I and T4D constructs was designated dvC904. For the *in vitro* experiments described in this chapter, we pre-determined an MOI of 200'000 vg/cell of each construct in the co-transduction paradigm, meaning a total of 400'000 vg/cell for dvC901 and dvC904. As a control virus, we produced large scale AAV PHP.eB from an identical construct but cloned with a control sgRNA, Ctrl3, designed to bind in an intergenic region ~12 Kb upstream of the

C9orf72 gene. Having no need to follow a co-transduction protocol for the control virus, a single MOI of 400'000 vg/cell was followed for dvCtrl3 in this chapter. Titration by qPCR confirmed that the large-scale productions were successful in producing AAV PHP.eB viruses with high titers (Table 6.1).

Table 6.1. AAV PHP.eB viral titers. List of viral titers of large-scale PHP.eB productions, expressed as viral genome copies per ml (vg/ml).

Virus	**Titer**
AAV PHP.eB CMV-GFP	$9.37x10^{11}$ vg/ml
AAV PHP.eB pX601 Ctrl3	$7.05x10^{12}$ vg/ml
AAV PHP.eB pX601 T4I	$1.06x10^{13}$ vg/ml
AAV PHP.eB pX601 T4D	$8.62x10^{12}$ vg/ml
AAV PHP.eB pX601 T5B	$2.25x10^{13}$ vg/ml

Having produced the viruses, we were then ready to test their transduction efficiency in C9-500 BAC cortical neurons. Immunocytochemistry on these cells showed that all viruses tested were able to express HA-tagged SaCas9 in these neurons, with a transduction efficiency ranging from 20%-30%, on average (Figure 6-2 A and B). Interestingly, although this efficiency is not dramatically different from that reported for our single-(viral)vector AAV9 (chapter 4.2), the fluorescence quantification revealed that, collectively, our new PHP.eB/CMV viruses tended to induce much stronger SaCas9 expression in transduced cells than the AAV9/EFS1a system, despite having used a lower MOI (Figure 6-2 C).

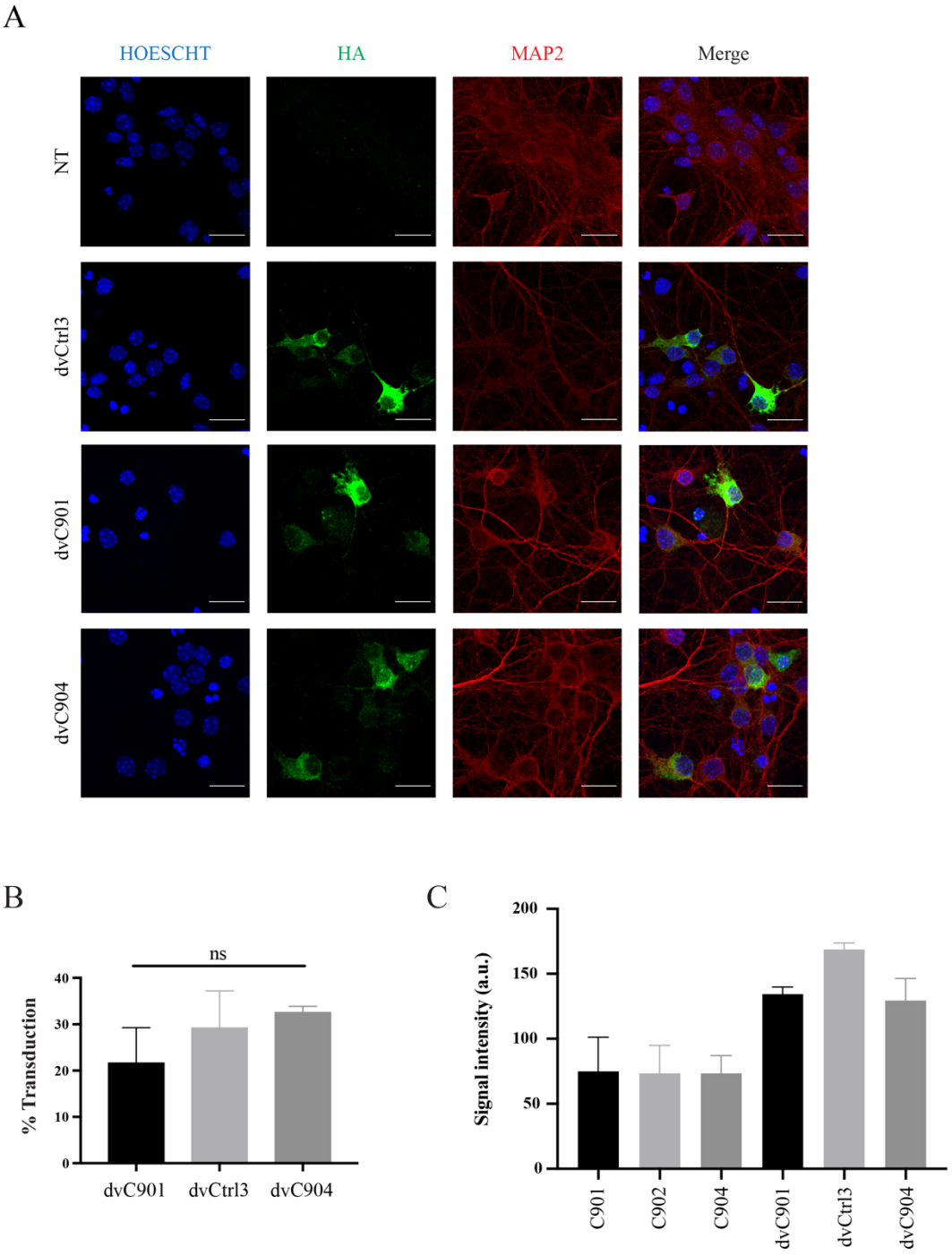


Figure 6-2. The dual-(viral)vector AAV PHP.eB/CMV system expresses HA-tagged Cas9 more efficiently than AAV9/EFS in C9-500 BAC cortical neurons. (A) Confocal microscope images showing HA expression upon 6-day transduction of C9-500 BAC cortical neurons with dvCtrl3, dvC901 and dvC904 (total of 400'000 vg/ cell). Scale bar = 20 µm. (B) Quantification of HA-positive neurons transduced with dvC901, dvCtrl3 and dvC904 for 6 days, showing no significant difference

between treated groups (n=3 biological replicates). One-way ANOVA, ns = not significant. (C) Quantification of HA signal intensity in transduced cells shows that PHP.eB/CMV system tends to induce considerably stronger expression in these neurons than AAV9/EFS1a, even with a lower MOI (n=3 biological replicates). a.u. = arbitrary units.

Having demonstrated their ability to express the transgene in target cells, it was now warranted to check whether dvC901 and dvC904 could mediate C9orf72 editing *in vitro*. With this objective in mind, we transduced C9-500 BAC cortical neurons with our viral vectors for a total of 6 days, after which gDNA was extracted from cells and screened for the presence of edited C9orf72 molecules. Using the method described in previous chapters, endpoint PCR across the G4C2 repeat region in non-treated (NT) or dvCtrl3-treated neurons did not amplify a product, whereas the same PCR in dvC901-treated cells amplified edited C9orf72 (Figure 6-3 A). As for dvC904, amplification of the region immediately upstream of the expansion yielded a smaller edited band in neurons treated with this therapy, indicating a targeted deletion took place (Figure 6-3 B). For each approach, neurons from two different embryos were used to ensure reproducibility of the results.

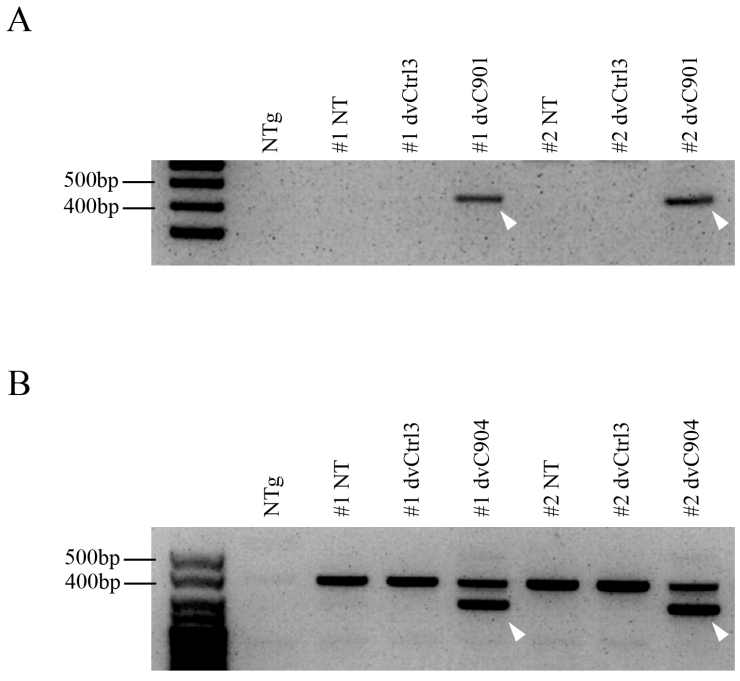


Figure 6-3. Transduction with dvC901 or dvC904 can edit the human C9orf72 gene in cortical neurons isolated from C9-500 BAC mice. (A) Agarose gel showing consistent excision of a full-size G4C2 expansion with dvC901 in C9-500 BAC cortical neurons. PCR amplification with C9_OUT_2 primers flanking the G4C2 repeats does not yield a product in non-treated (NT) or dvCtrl3-transduced neurons due to the GC-rich and repetitive nature of the expansion. In dvC901-treated neurons (total 400'000 vg/ cell), the excision of the expansion enables a ~400bp PCR product to form (white arrowheads). #1 and #2 represent cortical neurons from two different embryos. (B) Agarose gel showing targeted editing of human C9orf72 by dvC904 in C9-500 BAC cortical neurons. PCR amplification with C904_INT_14 primers amplifies an unedited product in all transgenic neurons. Transduction with dvC904 therapeutic virus (total 400'000 vg/ cell) introduces a targeted ~90bp deletion that is amplified by the same PCR (white arrowheads). #1 and #2 represent cortical neurons from two different embryos.

Having shown that dvC901 and dvC904 could edit the C9orf72 in C9-500 BAC neurons, we were ready to test whether these modified strategies would also decrease RNA Foci accumulation. To do this, we transduced transgenic cortical neurons with the AAV PHP.eB therapeutic and control viruses for a total of 6 days after which we did FISH staining on fixed cells. We then scanned the plates in an automated high-throughput microscope (Opera Phenix, Perkin Elmer) and used an unbiased automatic method to quantify FISH-positive inclusions. It was interesting to observe that dvC901 transduction significantly reduced the percentage of neurons positive for RNA Foci (Figure 6-4 A). Importantly, there was a decrease of ~40% and ~32% relative to non-treated (NT) and dvCtrl3-treated neurons, respectively, which is drastic when compared with equivalent results with the single-(viral)vector AAV9/EFS1a system (chapter 4.3). Considering that each RNA Foci is made up of a single expanded C9orf72 transcript (Liu et al., 2017), we can speculate that this sharper decrease by dvC901 in relation to its single-vector format, C901, was due to a more robust editing in the treated population. Interestingly, dvC904 showed a slight tendency to decrease Foci(+) neurons but this did not reach statistical significance, suggesting that, overall, the C904 strategy does not have an impact on RNA Foci accumulation.

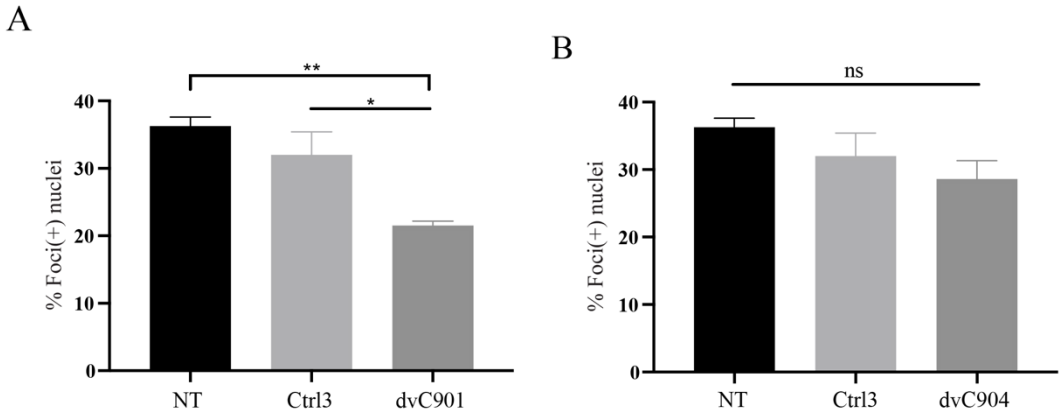


Figure 6-4. Transduction of C9-500 BAC cortical neurons with dvC901, but not dvC904, robustly decreases RNA Foci accumulation. High throughput quantification using Harmony software (Perkin Elmer) showed how transduction with (A) dvC901 for 6 days significantly reduced the percentage of neurons expressing sense RNA Foci by ~40% and ~32% when comparing with non-treated (NT) and dvCtrl3-treated neurons (n=5 biological replicates), respectively. (B) Transduction with dvC904 showed a tendency to reduce RNA Foci in treated neurons, but it did not reach statistical significance (n=5 biological replicates). One-way ANOVA, ns = not significant, *p<0.05, **p<0.01.

Having shown important *in vitro* evidence of a stronger, more robust effect of these dual-vector approaches, it was now important to check whether this enhanced impact would be replicated in an *in vivo* system. Therefore, as a preliminary pilot study, we injected one transgenic C9-500 BAC female mouse with dvC901 and one with dvC904, in the tail vein at 6 weeks of age. Similar to *in vitro* co-transductions, we used a co-injection protocol, containing $1x10^{11}$ vg of each composing virus, totalling $2x10^{11}$ vg per mouse. This dose was based on published guidelines (Chan et al., 2017) and corresponds to 20% of the viral dose used with the AAV9/EFS1a viruses described in the previous chapter. The mice were then kept for 6 weeks, during which time they did not show any signs of disease or distress. Upon sacrificing the mice and collecting tissue, we proceeded to analyse the extracted DNA by endpoint PCR with C9_OUT_2 and C904_INT_14 primers to detect DNA edited by dvC901 or dvC904, respectively, as described previously. We were excited to verify that intravenous injection with

dvC901 for 6 weeks was enough to edit the human C9orf72 gene throughout all the brain and spinal cord regions analysed (Figure 6-5 A). Similarly, the mouse treated with dvC904 had widespread editing in the spinal cord and the parietal-temporal lobes of the brain, although no edited C9orf72 was detected in the frontal lobe and the cerebellum (Figure 6-5 B). As discussed before, we suspect that the reason behind this is linked to the sensitivity of the detection method. It is important to note that this is early, preliminary data, corresponding to one mouse per treatment, and subsequent studies are needed to ensure reproducibility of these results. Our plans to increase the number of mice in this study were compromised by the COVID-19 crisis and will be re-evaluated in the near future.

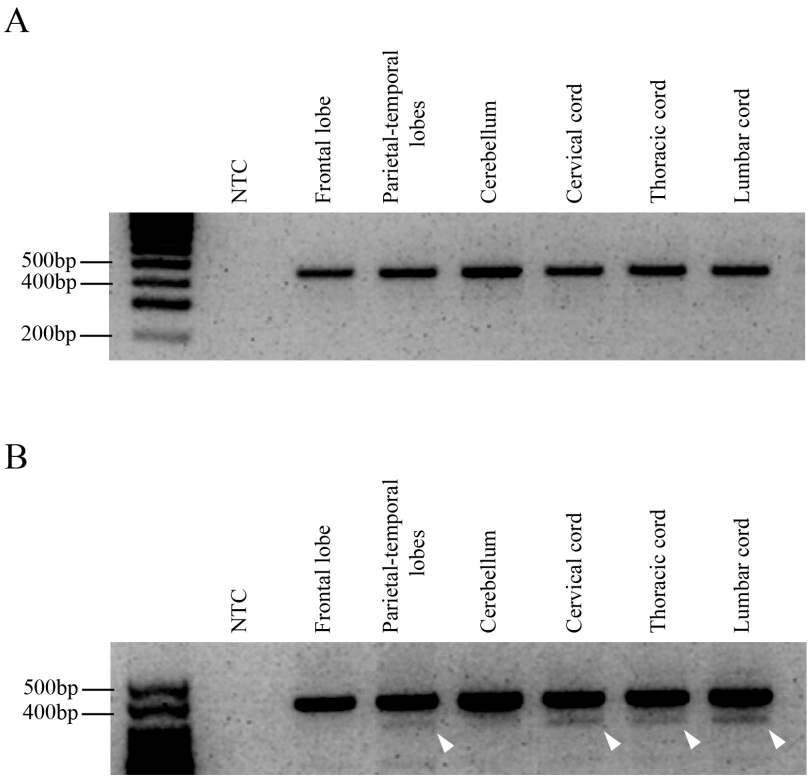


Figure 6-5. Intravenous injection of dvC901 and dvC904 can edit the human C9orf72 gene in widespread areas of the brain and spinal cord of C9-500 BAC mice. (A) Agarose gel showing widespread *in vivo* excision of the G4C2 expansion in the CNS of a C9-500 BAC mouse injected with dvC901 in the tail vein at 6 weeks of age. The animal was sacrificed 6 weeks post-injection. PCR amplification with C9_OUT_2 primers flanking the expansion was done on extracted

DNA, identifying edited C9orf72 (~400bp band) in all regions of the brain and spinal cord analyzed. (B) Agarose gel showing *in vivo* editing by dvC904 in several CNS regions of a C9-500 BAC mouse injected intravenously at 6 weeks of age and sacrificed 6 weeks post-injection. PCR amplification with C904_INT_14 primers allows detection of dvC904-edited C9orf72 in the parietal-temporal lobes of the brain, as well as cervical, thoracic and lumbar spinal cord (white arrowheads).

6.3 Discussion

In this chapter, we described modifications to our proposed therapeutic strategies, focused mainly on optimising viral delivery and the promoter-mediated expression. To this end, we decided to test a dual-vector approach which allowed us to introduce a larger and stronger promoter, and we also chose a different AAV capsid, PHP.eB, to deliver the therapeutic constructs. These changes were successful in mediating stronger transgene expression in cultured neurons, which resulted in a pronounced decrease in RNA Foci accumulation upon treatment with dvC901. We then provided preliminary data that shows how dvC901 and dvC904 are capable of widespread editing of the human C9orf72 gene in the CNS of treated mice.

The success of our modified approaches described in this chapter was surprising as much as it was exciting. Indeed, in the previous chapters, we had focused our efforts in designing and testing a single-vector editing approach over a dual-vector system, a direction we deemed more promising for future clinical translation due to the high costs of clinical-grade AAV manufacturing (Ayuso, 2016). In addition to this, we had hypothesised that a single-(viral)vector system would be more efficient, since co-transduction and co-injection approaches are dependent on each cell being transduced by two different viruses, which is statistically less likely than single transduction. Therefore, it was interesting to observe that, despite these facts, dvC901 and dvC904 not only retained their capability to edit the C9orf72 gene (Figure 6-3), but also seemed to mediate stronger transgene expression in cortical neurons than their single-vector formats (Figure 6-2), even with a lower MOI. In addition to this, dvC901 was considerably more efficient in reducing RNA Foci in these cells than what we had observed for C901 (chapter 4.3), suggesting that it edits the C9orf72 gene more efficiently as

well (Figure 6-4). Taking into account these experiments are in cultured cortical neurons, without BBB crossing requirement, it is likely that this increased impact we observed *in vitro* is more directly linked to the dual-vector configuration than the new capsid. These results confirmed that the inherent disadvantages of a system with two different viruses can in reality be overcome by the use of a strong promoter over the EFS1a, an observation which had been noted previously (Tabebordbar et al., 2016). In turn, the choice of the CMV instead of other promoters was down to its ubiquitous and well-described strong expression, making it an appropriate tool for a multi-system disease such as C9orf72-ALS/FTD (Grossman, 2019; Kim et al., 2013; Renton et al., 2014). It is important to note that the CMV is not ideal for gene replacement therapies due to its tendency to become methylated and silenced over time across different cell lines (Hsu et al., 2010; Osterlehner et al., 2011) but also in CNS tissue upon viral delivery (Brooks et al., 2004; Klein et al., 1998; McCown et al., 1996; Tenenbaum et al., 2004). However, we believe this issue would not constitute a disadvantage in the scope of our project. Indeed, for our proposed therapies, transgene expression only needs to be active long enough for DNA editing to take place, after which point the silencing of the promoter is not only not a problem, but it might even be desirable, potentially decreasing off-target cutting by SaCas9.

Considering the enhanced efficiency shown by the dual-vector system, it was therefore unexpected to observe that dvC904 did not significantly decrease RNA Foci in cortical neurons (Figure 6-4 B), despite our single-(viral)vector AAV9/C904 having done so (chapter 4.3). Taking into account the considerably sharper impact of dvC901 (Figure 6-4 A) on RNA Foci in comparison with AAV9/C901, we were expecting a similar tendency for dvC904. Taking these results into account, we could conclude that the C904/dvC904-mediated deletion does not decrease RNA Foci, and that the previously observed effect might have been the result of an indirect, non-specific effect from the sgRNAs used. However, we plan to explore this issue more in depth and to better understand the differences between PHP.eB/dual-vector and AAV9/single-vector by increasing the number of *in vitro* biological replicates and quantitatively comparing the editing efficiencies of each system.

In the last experiments of this chapter, we describe a preliminary set of data concerning the *in vivo* potential of dvC901 and dvC904. This data needs to be interpreted with significant caution due to the reduced number of animals involved (n=1 per treated group), but we deemed it worthwhile to report due to its exciting implications. Indeed, an intravenous injection of dvC901 in a 6 week-old mouse resulted in widespread excision of a pathogenic G4C2 expansion from the human C9orf72 gene in every region of the CNS analysed (Figure 6-5 A).

This result shows a significantly increased widespread editing in the CNS when compared with AAV9/C901, and is even more remarkable considering we used 80% less viral load than what we used for the *in vivo* AAV9/EFS1a study (chapter 5.3). Similarly, a mouse injected with dvC904 displayed targeted C9orf72 editing throughout the entire spinal cord, as well as the parietal-temporal lobes of the brain (Figure 6-5 B), a dramatic improvement when compared to its single-vector format (chapter 5.3). In the near future, we plan to increase the number of mice treated with each therapy and quantify the *in vivo* editing efficiency, after which we plan to study the impact of these optimised approaches in the poly-(GP) accumulation in the brain and spinal cord.

Despite the exciting data, there are important considerations about the use of AAV PHP.eB, and whether it would be translatable into clinical use. Indeed, as mentioned above, there were concerning reports about its parent capsid, PHP.B, indicating that the striking neurotropism observed was limited to only some mouse strains including C57BL/6, FVB/N, DBA/2 and SJL/J (Hordeaux et al., 2018; Matsuzaki et al., 2019). During the course of the experiments described in this chapter, additional work identified a GPI-anchored protein, LY6A, as a key ligand for both the PHP.B and PHP.eB capsids, responsible for its BBB permeability (Batista et al., 2020; Hordeaux et al., 2019; Huang et al., 2019). Importantly, primates lack a LY6A homolog (Loughner et al., 2016), possibly explaining why AAV PHP.B was unable to efficiently cross the BBB in non-human primates, reported in two separate studies (Hordeaux et al., 2018; Matsuzaki et al., 2018). This evidence forces us to assume that AAV PHP.eB, which was used to great success in this and other projects (Chan et al., 2017), is very likely to not be applicable for future use in humans. Nevertheless, the proof-of-concept described here is important to demonstrate how the therapeutic strategies proposed by us are scalable and might one day, with the development of new clinically promising capsids, lead to an elegant and ground-breaking therapy to treat C9orf72-linked neurodegeneration. Recently published work suggests this reality might even be closer than expected. Indeed, Kumar and colleagues (Ravindra Kumar et al., 2020) expanded on their previous work to develop a Multiplexed-CREATE system, from which they selected new PHP AAV capsid variants, named PHP.C, with enhanced neurotropism that are not background strain-specific. This suggests that the new PHP.C variants might use different receptors other than LY6A to cross the BBB and might therefore have increased clinical relevance.

In summary, this chapter describes our strategy to optimise the delivery and expression of the editing approaches we had designed, with the ultimate goal of improving their therapeutic

impact. This strategy was built on data from previous chapters as well as the literature, leading us to test a CMV-mediated dual-vector configuration delivered by the highly neurotropic AAV capsid PHP.eB. These optimised approaches were able to edit the human C9orf72 gene and induce stronger transgene expression in cultured cortical neurons, while dvC901 dramatically decreased RNA Foci-expressing neurons *in vitro*. In addition to this, we show preliminary *in vivo* evidence for widespread editing of human C9orf72 in the CNS of C9-500 BAC mice following a one-time intravenous injection, although further *in vivo* investigations are needed to complete this study. These results put together offer a glimpse of what a revolutionary treatment for C9orf72-ALS/FTD could look like in the not too distant future, and they open the door for more relentless research to make this a reality for patients affected by these truly devastating disorders.

7. Final discussion and future perspectives

7.1 Summary and project highlights

ALS and FTD are two fatal diseases that relentlessly and progressively disable patients, in most cases leading to death at 5 years or 6 to 11 years after diagnosis, respectively (Bang et al., 2015; Hodges et al., 2003; Talbot, 2009). In addition to the primary and more obvious neurodegenerative symptoms, research into the psychological burden on patients and caregivers as well as the economic impact on affected families reveals the truly devastating nature of these disorders. Indeed, heart-breaking studies have disclosed how ALS patients and their caregivers are afflicted by hopelessness, end-of-life anxiety and depression, in addition to staggering out-of-pocket costs related to disease-associated home renovations, mobility aids, medical expenses, among many others (Armon, 2006; Averill et al., 2007; Gauthier et al., 2007; Gladman et al., 2014; Grabler et al., 2018; Mockford et al., 2006). Similarly, the neuropsychiatric and behavioural abnormalities suffered by FTD patients (Bang et al., 2015; Olney et al., 2017) are associated with caregivers' distress and poor quality of life (Hvidsten et al., 2019; Mourik et al., 2004; Riedijk et al., 2006), while the economic burden on affected families is estimated to be greater than that of Alzheimer's due to higher loss-of-productivity costs (Galvin et al., 2017). Despite these devastating consequences, the available treatments for ALS and FTD are still tragically inefficient in significantly slowing disease progression and improving the survivability of patients. This reality highlights the urgent need for progress in the treatment of these disorders, which can only be achieved by relentlessly researching their causes and pathophysiology, and using this knowledge to continuously develop new therapeutic avenues. Being the major genetic origin so far discovered for both diseases, the hexanucleotide expansion of the C9orf72 gene is an attractive therapeutic target for which an effective treatment would benefit a large number of patients and could have an enormous impact on affected families.

In this thesis, we have explored a role for the revolutionising gene editing technology CRISPR-Cas9 as a therapeutic agent to treat C9orf72-ALS/FTD, having produced novel scientific knowledge and developed the very first constructs that might one day lead to a fully translated therapy. We devised two new bold approaches, C901 and C904, that either excise the full pathogenic expansion, or a ~90 bp sequence containing an alternative Kozak consensus thought to drive RAN translation of the expansion, respectively. In chapter 3, we describe the design of the sgRNAs and their cloning into a dual-vector or a cloned single-vector configuration,

after which we reported how these constructs can introduce the desired deletions with high efficiency in a human cell line with 3xG4C2 repeats. Importantly, this editing was found to not impact on endogenous C9orf72 expression. We then packaged our cloned single-vector constructs in the therapeutically relevant AAV9 viral vector and used them to treat primary neurons expressing a pathogenic C9orf72 allele derived from a human patient (chapter 4). Interestingly, AAV9-delivered C901 and C904 successfully introduced the desired deletions in this allele in a reproducible manner, reducing the molecular by-products of the expansion as a consequence. We then proceeded to validate our therapeutic strategies in adult C9-500 BAC transgenics (Liu et al., 2016). The behavioural characterisation described in chapter 5 provides novel and crucial insights into the C9-500 BAC model that might inform other teams planning to do basic and translational research with these mice. Namely, we report an FVB/N background-associated severe syndrome that seriously interferes with proper behavioural analysis and could have been caused by backcrossing transgenic animals to an FVB/N colony with a strong epileptic predisposition. However, this model has a strong genetic and molecular basis, having been created with a full-length human expanded C9orf72 allele including introns and flanking regions and expressing G4C2 expansion by-products. This makes it an attractive model for basic molecular and gene therapy proof-of-concept studies. Taking advantage of this, we were, to the best of our knowledge, the first to show excision of a pathogenic human C9orf72 expansion in the CNS following a one-time intravenous administration in mice with an AAV9-delivered gene therapy (Figure 7-1). Importantly, animals injected with this experimental treatment had decreased accumulation of poly-(GP) DPRs in the brain, an encouraging result considering the promising characteristics of poly-(GP) as a pharmacodynamic marker with potential to be used in future C9orf72 clinical trials (Balendra et al., 2017; Gendron et al., 2017; Lehmer et al., 2017) (Figure 7-1). Finally, in an important step towards future clinical translation, we also demonstrated how the efficiency of our therapeutic approaches can be scaled up (chapter 6). This suggests that future modifications to the viral vectors, the transgenes and the delivery routes might produce significantly better outcomes. Indeed, we report how the use of a dual-vector approach with a strong promoter and a capsid with increased neurotropism can result in significant reduction of RNA Foci in cultured neurons and exciting widespread C9orf72 editing in the CNS of treated mice (Figure 7-1). The data presented here invites for further research and optimisations, and proposes a future where patients with C9orf72-ALS/FTD might one day be offered the therapeutic option of having their pathogenic mutation partially corrected.

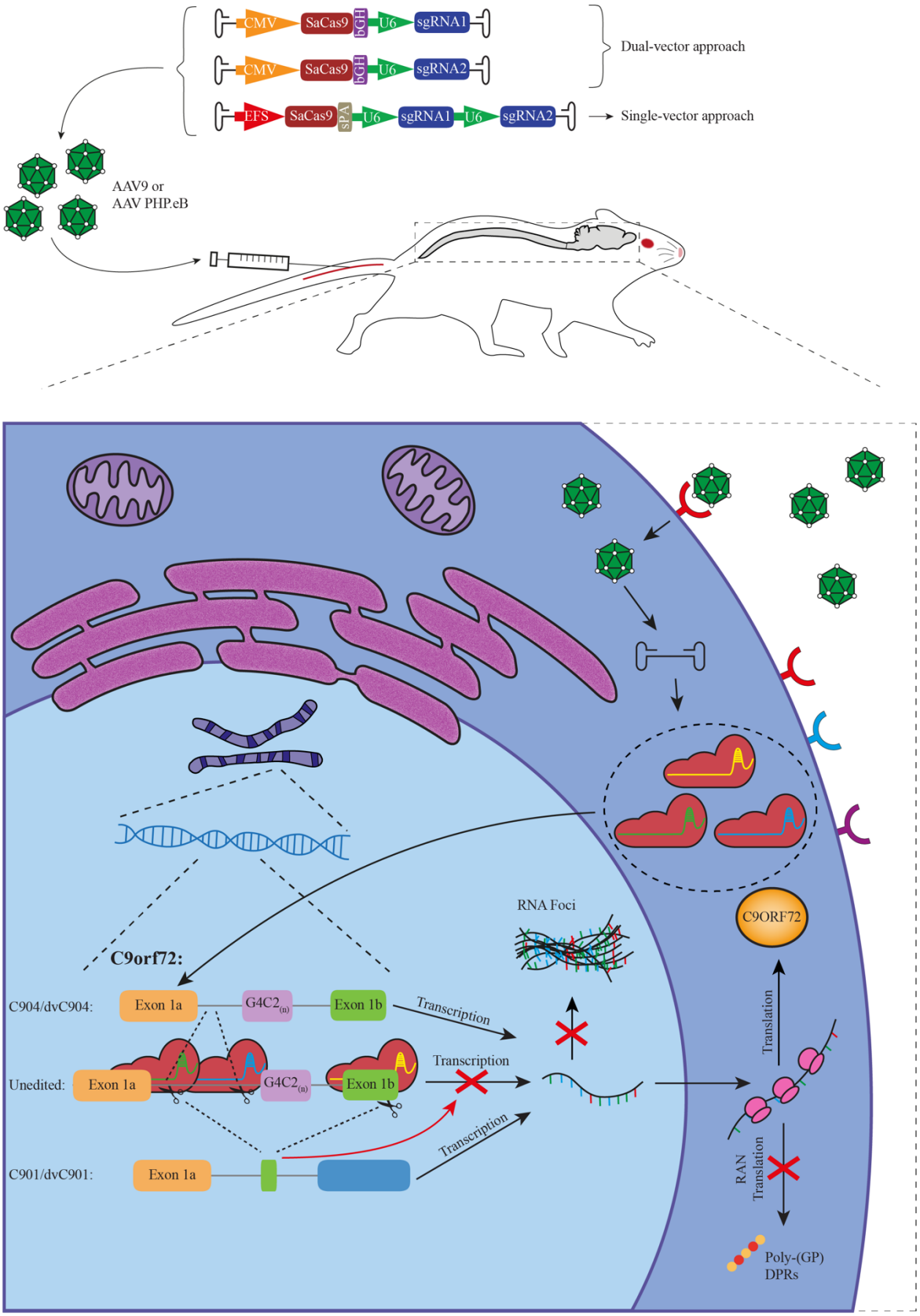

CMV
SaCas9
bGH
U6
sgRNA1
CMV
SaCas9
bGH
U6
sgRNA2
Dual-vector approach
EFS
SaCas9
sPA
U6
sgRNA1
U6
sgRNA2
Single-vector approach
AAV9 or
AAV PHP.eB
C9orf72:
C904/dvC904:
Exon 1a
G4C2(n)
Exon 1b
Transcription
Unedited:
Exon 1a
G4C2(n)
Exon 1b
Transcription
C901/dvC901:
Exon 1a
Transcription
RNA Foci
C9ORF72
Translation
RAN
Translation
Poly-(GP)
DPRs

Figure 7-1. Project outcomes and highlights: the promise of therapeutic *in vivo* editing of human expanded C9orf72. Schematic illustration summarising the main outcomes of this project with a focus on the *in vivo* therapeutic potential. The C901/C904 or dvC901/dvC904 experimental therapies rely on using either a single-vector or a dual-vector approach, respectively, to express SaCas9 with a pair of sgRNAs that introduce strategic deletions in the human C9orf72 gene in a programmable manner. These constructs can be packaged to form recombinant AAV9 or AAV PHP.eB viruses that can transduce neurons *in vitro* but can also be administered systemically *in vivo* through injection in the tail vein of mice. Following a one-time intravenous delivery in C9-500 BAC transgenics, AAVs can cross the BBB and transduce different regions of the CNS. Upon transduction of CNS cells, the transgene is unpacked from the viral capsid and expresses SaCas9 protein complexed with therapeutic sgRNAs. The SaCas9/sgRNA complexes can then enter the nucleus and edit human C9orf72 DNA, either excising the entire $G4C2_{(n)}$ expansion or a highly regulatory region located upstream of the repeats, finally promoting the re-ligation of the gene across the cut sites. In edited cells, C901-mediated excision of the expansion stops transcription of unedited C9orf72 and consequent accumulation of RNA Foci, as well as poly-(GP) DPRs. From experiments in HEK293T cells with robust G4C2 excision (chapter 3.8), this editing is unlikely to interfere with endogenous C9orf72 expression. Editing with C904 strategy was only observed in the CNS by using the PHP.eB/CMV system (dvC904), and further research is needed to ascertain whether deleting the targeted ~90 bp sequence has therapeutic potential. Treatment with dvC901 and dvC904 results in more widespread CNS editing *in vivo* due to the enhanced neurotropism of AAV PHP.eB and the use of the stronger CMV promoter (chapter 6).

Interestingly, the G4C2 expansion of the C9orf72 can also be detected in patients with other neurodegenerative traits including AD, PLS and Parkinsonian phenotypes, among others, although the causality of this association is still not entirely clear (Cooper-Knock et al., 2012, 2014b). This suggests that a future therapy like the one we explored in this thesis might have broader applicability than purely ALS/FTD cases.

7.2 Main challenges and future directions

The difficulty with the behavioural assessment of C9-500 BAC mice was perhaps the biggest and most surprising challenge of this project. Based on our data and interpretation of the literature, we believe important work lies ahead to uncover the mystery of the background-associated epileptic syndrome in this model and its confounding effects on possible C9orf72-related symptoms. As discussed in chapter 5, a possible explanation to the discrepancies reported might relate to the backcrossing of the donated C9-500 BAC transgenics to an inbred FVB/N colony with a strong epileptic predisposition. In the future, a possible solution to this issue would be to establish a completely new colony using non-transgenic and transgenic mice from the original source (Liu et al., 2016) and maintaining it by exclusively crossing littermates. As an alternative, we could proceed to transfer the C9orf72 transgene in our colony into a distinct genetic background without neurological traits by continuous backcrossing, a procedure routinely done in mouse genetics for over two decades to develop congenic models (Montagutelli, 2000). This, however, would not be a risk-free approach. Indeed, even after 10 generations of backcrossing and transgene selection, it is common for congenic mice to still have passenger DNA from the original strain, usually corresponding to sequences flanking the transgene insert (Vanden Berghe et al., 2015; Lusis et al., 2007). Concerning the C9-500 BAC, the genetic information responsible for the background-associated neurological syndrome could, albeit unlikely, be contained within this passenger DNA. In this scenario, the consequences would be disastrous, as there would be no appropriate controls, leaving no way of determining whether symptoms in transgenic animals originated from C9orf72-linked neurodegeneration or unrelated consequences from epilepsy. Considering all this, the safest option, and also the most resource-consuming one, would be to use the same BAC DNA and insert it directly into the genome of a more reliable target background, thereby avoiding subsequent backcrossing with different strains (Vanden Berghe et al., 2015). Overall, a reliable C9orf72 transgenic mouse model with a clear ALS/FTD-like phenotype would be a great asset to investigate the effect of this and other therapeutic strategies on disease symptoms, and one of the approaches proposed above could achieve this goal in the future. While these experiments proceed, however, the genetic design of the C9-500 BAC mice and its molecular characteristics make it an appealing model to continue optimising and upgrading the therapeutic strategies described in this thesis.

An important aspect that warrants such optimisation is the delivery route of our viral vectors. In chapter 6, we described exciting preliminary evidence for widespread C9orf72 therapeutic editing in the CNS of treated mice following intravenous administration, which could constitute a revolutionary non-invasive treatment for C9orf72-ALS/FTD. However, recent work published after the start of our experiments reported that the remarkable neurotropic properties of the PHP.eB capsid are limited to animals expressing the GPI-anchored protein LY6A in the brain endothelial cells forming the BBB. Primates lack a homolog for this ligand, rendering the enhanced neurotropism of PHP.eB not translatable to humans (Batista et al., 2020; Hordeaux et al., 2019; Huang et al., 2019; Loughner et al., 2016). Recently, new AAV PHP capsid variants were described whose enhanced neurotropism is not limited to mouse strains expressing LY6A, raising hopes that these results might be reproducible in primates. If this turns out to be the case, more experiments will need to be done with our own constructs and a non-invasive, intravenously administered treatment might still be possible to develop. A suitable alternative to intravenous administration is the direct delivery of the treatment into the CSF, either by intrathecal or intraventricular injection, bypassing the BBB altogether (S. Hersh et al., 2016). Recently, intrathecal administration has been successfully used in humans to deliver gene therapies for SOD1-ALS and was regarded relatively safe, with one of these studies even reporting a reduction of SOD1 protein concentrations in the CSF upon ASO therapy (Miller et al., 2020; Mueller et al., 2020). In mice, intracisternal injection can also be used to efficiently deliver gene therapies into the CSF (Iannitti et al., 2018; Lukashchuk et al., 2016) but the technically challenging nature of this procedure led us to initially opt by tail vein administration. Nevertheless, after showing that a more efficient viral vector delivery remarkably increased C9orf72 editing in the CNS (chapter 6) and considering the exciting recent results from clinical trials (Miller et al., 2020), we will also explore this delivery route for our therapeutic strategies in the near future.

Although falling outside the scope of this project, whose initial objective focuses on developing and optimising the therapeutic potential of our strategies, a thorough study of off-target cuts arising from our approaches will have to be performed in the future and before clinical testing. This stems from the concern that the therapeutic benefit of editing expanded C9orf72 could be counterbalanced by rogue, off-target activity by Cas9, resulting in the accumulation of off-target dsDNA breaks, some of which might incur in secondary gene disruption and toxicity. Detecting off-target cuts by Cas9 is a more complex task than detecting on-target ones because the frequency and location of off-targets in the genome are not known *a priori*. Nevertheless,

several methods have been developed so far and the majority of them have been extensively reviewed elsewhere (Zhang et al., 2015; Zischewski et al., 2017). Briefly, these methods can consist of biased approaches, in which off-target sites are computationally predicted and can then be further amplified and confirmed by sequencing (Cho et al., 2014; Ran et al., 2013a), or unbiased approaches that provide a genome-wide perspective of Cas9-induced dsDNA breaks. Biased methods have the advantages of being affordable for most labs and would provide an initial indication of the frequency of off-target mutations mediated by our experimental therapies. Following this initial screening, however, we aim to apply a genome-wide, unbiased technique, among which digested genomes (Digenome)-seq stands out as a robust and cost-effective method, widely adopted by several research groups (Kim et al., 2015; Zhang et al., 2015). Indeed, Digenome-seq involves the *in vitro* digestion of cell-free genomic DNA with Cas9 and the sgRNAs to be screened, as well as a mock-digested genome, followed by whole-genome sequencing (WGS) of the products. Then, a freely-accessible web-based tool (Park et al., 2017) can identify and count sequencing reads with identical 5' ends, which indicates a dsDNA blunt cut has occurred. Importantly, because the digestion is made in cell-free genomic DNA, off-targets are not repaired by the cell, resulting in a more comprehensive perspective of the off-target activity and, therefore, potential safety of our experimental therapies. Further to the approaches mentioned above, an additional study might also be necessary to investigate off-target effects of our CRISPR therapies in an *in vivo* context. The recently published detection method 'discovery of in situ Cas off-targets and verification by sequencing' (DISCOVER-Seq) is an attractive technique that we could use for this purpose (Wienert et al., 2019). Harnessing the endogenous DNA repair mechanism, DISCOVER-Seq works by isolating the meiotic recombination 11 (MRE11) protein, which identifies Cas9-induced DNA breaks, and sequencing bound DNA (Wienert et al., 2019). Therefore, this detection method bypasses the need for external reagents and allows an unbiased analysis of Cas9-induced off-targets in a cellular or tissue context, complementing the cell-free approach of Digenome-seq. It is important to note that, although precautions must and will be taken, off-target cuts are actually considerably less frequent than on-target ones in cells with fully-functioning DNA repair mechanisms (Zischewski et al., 2017). In addition to this, SaCas9 was expressed in C9-500 BAC mice for over one year without any visible consequences or obvious toxicity in the treated groups and, although not specific to editing, C901- and C904-treated mice even had a tendency for slightly better survival (chapter 5). This gives us confidence that off-targets from our treatments are probably negligible and might not pose a serious safety concern, although appropriate toxicology and safety studies will need to be conducted.

8. References

Aartsma-Rus, A., and van Putten, M. (2014). Assessing functional performance in the mdx mouse model. J. Vis. Exp. *85*, 51303.

Ababneh, N.A., Scaber, J., Flynn, R., Douglas, A., Barbagallo, P., Candalija, A., Turner, M.R., Sims, D., Dafinca, R., Cowley, S.A., et al. (2020). Correction of amyotrophic lateral sclerosis related phenotypes in induced pluripotent stem cell-derived motor neurons carrying a hexanucleotide expansion mutation in C9orf72 by CRISPR/Cas9 genome editing using homology-directed repair. Hum. Mol. Genet. *29*, 2200–2217.

Abe, K., Aoki, M., Tsuji, S., Itoyama, Y., Sobue, G., Togo, M., Hamada, C., Tanaka, M., Akimoto, M., Nakamura, K., et al. (2017). Safety and efficacy of edaravone in well defined patients with amyotrophic lateral sclerosis: a randomised, double-blind, placebo-controlled trial. Lancet Neurol. *16*, 505–512.

Abo-Rady, M., Kalmbach, N., Pal, A., Schludi, C., Janosch, A., Richter, T., Freitag, P., Bickle, M., Kahlert, A.K., Petri, S., et al. (2020). Knocking out C9ORF72 Exacerbates Axonal Trafficking Defects Associated with Hexanucleotide Repeat Expansion and Reduces Levels of Heat Shock Proteins. Stem Cell Reports *14*, 390–405.

Abrahams, S., Goldstein, L.H., Kew, J.J., Brooks, D.J., Lloyd, C.M., Frith, C.D., and Leigh, P.N. (1996). Frontal lobe dysfunction in amyotrophic lateral sclerosis. A PET study. Brain 2105–2120.

Al-Chalabi, A., and Hardiman, O. (2013). The epidemiology of ALS: a conspiracy of genes, environment and time. Nat. Rev. Neurol. *9*, 617–628.

Al-Chalabi, A., Hardiman, O., Kiernan, M.C., Chiò, A., Rix-Brooks, B., and van den Berg, L.H. (2016). Amyotrophic lateral sclerosis: moving towards a new classification system. Lancet Neurol. *15*, 1182–1194.

Almeida, S., Gascon, E., Tran, H., Chou, H.J., Gendron, T.F., DeGroot, S., Tapper, A.R., Sellier, C., Charlet-Berguerand, N., Karydas, A., et al. (2013). Modeling key pathological features of frontotemporal dementia with C9ORF72 repeat expansion in iPSC-derived human neurons. Acta Neuropathol. *126*, 385–399.

Alonso, A., Logroscino, G., Jick, S.S., and Hernán, M.A. (2009). Incidence and lifetime risk of motor neuron disease in the United Kingdom: a population-based study. Eur. J. Neurol. *16*, 745–751.

Altshuler, D.L., Durbin, R.M., Abecasis, G.R., Bentley, D.R., Chakravarti, A., Clark, A.G., Collins, F.S., De La Vega, F.M., Donnelly, P., Egholm, M., et al. (2010). A map of human genome variation from population-scale sequencing. Nature *467*, 1061–1073.

Anguela, X.M., and High, K.A. (2019). Entering the Modern Era of Gene Therapy. Annu. Rev. Med. *70*, 273–288.

Annegers, J.F., Appel, S., Lee, J.R., and Perkins, P. (1991). Incidence and prevalence of amyotrophic lateral sclerosis in Harris County, Texas, 1985-1988. Arch. Neurol. *48*, 589–593.

Anzalone, A. V., Randolph, P.B., Davis, J.R., Sousa, A.A., Koblan, L.W., Levy, J.M., Chen, P.J., Wilson, C., Newby, G.A., Raguram, A., et al. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. Nature *576*, 149–157.

Aoki, Y., Manzano, R., Lee, Y., Dafinca, R., Aoki, M., Douglas, A.G.L., Varela, M.A., Sathyaprakash, C., Scaber, J., Barbagallo, P., et al. (2017). C9orf72 and RAB7L1 regulate vesicle trafficking in amyotrophic lateral sclerosis and frontotemporal dementia. Brain *140*, 887–897.

Arai, T., Hasegawa, M., Akiyama, H., Ikeda, K., Nonaka, T., Mori, H., Mann, D., Tsuchiya, K., Yoshida, M., Hashizume, Y., et al. (2006). TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Biochem. Biophys. Res. Commun. *351*, 602–611.

Armon, C. (2003). An evidence-based medicine approach to the evaluation of the role of exogenous risk factors in sporadic amyotrophic lateral sclerosis. Neuroepidemiology *22*, 217–228.

Armon, C. (2006). Who cares for the carers? Amyotroph. Lateral Scler. *7*, 131.

Armon, C. (2009). Smoking may be considered an established risk factor for sporadic ALS. Neurology *73*, 1693–1698.

Arnold, C. (2019). Tailored treatment for ALS poised to move ahead. Nat. Med.

Arrant, A.E., Filiano, A.J., Unger, D.E., Young, A.H., and Roberson, E.D. (2017). Restoring neuronal progranulin reverses deficits in a mouse model of frontotemporal dementia. Brain *140*, 1447–1465.

Arthur, K.C., Calvo, A., Price, T.R., Geiger, J.T., Chiò, A., and Traynor, B.J. (2016). Projected increase in amyotrophic lateral sclerosis from 2015 to 2040. Nat. Commun. *7*, 12408.

Ash, P.E.A., Bieniek, K.F., Gendron, T.F., Caulfield, T., Lin, W.-L., DeJesus-Hernandez, M., van Blitterswijk, M.M., Jansen-West, K., Paul, J.W., Rademakers, R., et al. (2013). Unconventional Translation of C9ORF72 GGGGCC Expansion Generates Insoluble Polypeptides Specific to c9FTD/ALS. Neuron *77*, 639–646.

Atanasio, A., Decman, V., White, D., Ramos, M., Ikiz, B., Lee, H.-C., Siao, C.-J., Brydges, S., LaRosa, E., Bai, Y., et al. (2016). C9orf72 ablation causes immune dysregulation characterized by leukocyte expansion, autoantibody production, and glomerulonephropathy in mice. Sci. Rep. *6*, 23204.

Auton, A., Abecasis, G.R., Altshuler, D.M., Durbin, R.M., Bentley, D.R., Chakravarti, A., Clark, A.G., Donnelly, P., Eichler, E.E., Flicek, P., et al. (2015). A global reference for human genetic variation. Nature *526*, 68–74.

Averill, A.J., Kasarskis, E.J., and Segerstrom, S.C. (2007). Psychological health in patients with amyotrophic lateral sclerosis. Amyotroph. Lateral Scler. *8*, 243–254.

Ayuso, E. (2016). Manufacturing of recombinant adeno-associated viral vectors: new technologies are welcome. Mol. Ther. - Methods Clin. Dev. *3*, 15049.

Azzoni, A.R., Ribeiro, S.C., Monteiro, G.A., and Prazeres, D.M.F. (2007). The impact of polyadenylation signals on plasmid nuclease-resistance and transgene expression. J. Gene Med. *9*, 392–402.

Azzouz, M., Hottinger, A., Paterna, J.C., Zurn, A.D., Aebischer, P., and Büeler, H. (2000). Increased motoneuron survival and improved neuromuscular function in transgenic ALS mice after intraspinal injection of an adeno-associated virus encoding Bcl-2. Hum. Mol. Genet. *9*, 803–811.

Azzouz, M., Ralph, G.S., Storkebaum, E., Walmsley, L.E., Mitrophanous, K.A., Kingsman, S.M., Carmeliet, P., and Mazarakis, N.D. (2004). VEGF delivery with retrogradely transported lentivector prolongs survival in a mouse ALS model. Nature *429*, 413–417.

Bae, S., Park, J., and Kim, J.-S. (2014). Cas-OFFinder: a fast and versatile algorithm that

searches for potential off-target sites of Cas9 RNA-guided endonucleases. *30*, 1473–1475.

Baker, M., Mackenzie, I.R., Pickering-Brown, S.M., Gass, J., Rademakers, R., Lindholm, C., Snowden, J., Adamson, J., Sadovnick, A.D., Rollinson, S., et al. (2006). Mutations in progranulin cause tau-negative frontotemporal dementia linked to chromosome 17. Nature *442*, 916–919.

Balakrishnan, B., and Jayandharan, G.R. (2014). Basic biology of adeno-associated virus (AAV) vectors used in gene therapy. Curr. Gene Ther. *14*, 86–100.

Balendra, R., and Isaacs, A.M. (2018). C9orf72-mediated ALS and FTD: multiple pathways to disease. Nat. Rev. Neurol. *14*, 544–558.

Balendra, R., Moens, T.G., and Isaacs, A.M. (2017). Specific biomarkers for C9orf72 FTD / ALS could expedite the journey towards effective therapies . EMBO Mol. Med. *9*, 853–855.

Bang, J., Spina, S., and Miller, B.L. (2015). Frontotemporal dementia. Lancet *386*, 1672–1682.

Bannwarth, S., Ait-El-Mkadem, S., Chaussenot, A., Genin, E., Lacas-Gervais, S., Fragaki, K., Berg-Alonso, L., Kageyama, Y., Serre, V., Moore, D., et al. (2014). A mitochondrial origin for frontotemporal dementia and amyotrophic lateral sclerosis through CHCHD10 involvement. Brain *137*, 2329–2345.

Batista, A.R., King, O.D., Reardon, C.P., Davis, C., Shankaracharya, Philip, V., Gray-Edwards, H., Aronin, N., Lutz, C., Landers, J., et al. (2020). Ly6a Differential Expression in Blood-Brain Barrier Is Responsible for Strain Specific Central Nervous System Transduction Profile of AAV-PHP.B. Hum. Gene Ther. *31*, 90–102.

Batra, R., Nelles, D.A., Pirie, E., Blue, S.M., Marina, R.J., Wang, H., Chaim, I.A., Thomas, J.D., Zhang, N., Nguyen, V., et al. (2017). Elimination of Toxic Microsatellite Repeat Expansion RNA by RNA-Targeting Cas9. Cell *170*, 899-912.e10.

Beck, J., Poulter, M., Hensman, D., Rohrer, J.D., Mahoney, C.J., Adamson, G., Campbell, T., Uphill, J., Borg, A., Fratta, P., et al. (2013). Large C9orf72 hexanucleotide repeat expansions are seen in multiple neurodegenerative syndromes and are more frequent than expected in the UK population. Am. J. Hum. Genet. *92*, 345–353.

Becker, L.A., Huang, B., Bieri, G., Ma, R., Knowles, D.A., Jafar-Nejad, P., Messing, J., Kim,

H.J., Soriano, A., Auburger, G., et al. (2017). Therapeutic reduction of ataxin-2 extends lifespan and reduces pathology in TDP-43 mice. Nature *544*, 367–371.

Benchling (2020). [Biology software]. Retrieved from https://benchling.com.

Benjaminsen, E., Alstadhaug, K.B., Gulsvik, M., Baloch, F.K., and Odeh, F. (2018). Amyotrophic lateral sclerosis in Nordland county, Norway, 2000–2015: prevalence, incidence, and clinical features. Amyotroph. Lateral Scler. Front. Degener. *19*, 522–527.

Bensimon, G., Lacomblez, L., and Meininger, V. (1994). A Controlled Trial of Riluzole in Amyotrophic Lateral Sclerosis. N. Engl. J. Med. *330*, 585–591.

Vanden Berghe, T., Hulpiau, P., Martens, L., Vandenbroucke, R.E., Van Wonterghem, E., Perry, S.W., Bruggeman, I., Divert, T., Choi, S.M., Vuylsteke, M., et al. (2015). Passenger Mutations Confound Interpretation Of All Genetically Modified Congenic Mice. Immunity *43*, 200–209.

Bernardi, L., Frangipane, F., Smirne, N., Colao, R., Puccio, G., Curcio, S.A.M., Mirabelli, M., Maletta, R., Anfossi, M., Gallo, M., et al. (2012). Epidemiology and genetics of frontotemporal dementia: A door-to-door survey in Southern Italy. Neurobiol. Aging *33*, 2948.e1-2948.e10.

van Blitterswijk, M., DeJesus-Hernandez, M., and Rademakers, R. (2012). How do C9ORF72 repeat expansions cause amyotrophic lateral sclerosis and frontotemporal dementia. Curr. Opin. Neurol. *25*, 689–700.

van Blitterswijk, M., DeJesus-Hernandez, M., Niemantsverdriet, E., Murray, M.E., Heckman, M.G., Diehl, N.N., Brown, P.H., Baker, M.C., Finch, N.C.A., Bauer, P.O., et al. (2013). Association between repeat sizes and clinical and pathological characteristics in carriers of C9ORF72 repeat expansions (Xpansize-72): A cross-sectional cohort study. Lancet Neurol. *12*, 978–988.

Van Blitterswijk, M., Van Es, M.A., Hennekam, E.A.M., Dooijes, D., Van Rheenen, W., Medic, J., Bourque, P.R., Schelhaas, H.J., Van Der Kooi, A.J., De Visser, M., et al. Evidence for an oligogenic basis of amyotrophic lateral sclerosis.

Boivin, M., Pfister, V., Gaucherot, A., Ruffenach, F., Negroni, L., Sellier, C., and Charlet-Berguerand, N. (2020). Reduced autophagy upon C9ORF72 loss synergizes with dipeptide

repeat protein toxicity in G4C2 repeat expansion disorders. EMBO J. *39*.

Bonaparte, J.P., Grant, I.A., Benstead, T.J., Murray, T.J., and Smith, M. (2007). ALS incidence in Nova Scotia over a 20-year-period: a prospective study. Can. J. Neurol. Sci. *34*, 69–73.

Borel, F., Gernoux, G., Cardozo, B., Metterville, J.P., Toro Cabreja, G.C., Song, L., Su, Q., Gao, G.P., ElMallah, M.K., Brown, R.H., et al. (2016). Therapeutic rAAVrh10 Mediated SOD1 Silencing in Adult SOD1G93A Mice and Nonhuman Primates. Hum. Gene Ther. *27*, 19–31.

Brooks, B. (1994). El Escorial World Federation of Neurology criteria for the diagnosis of amyotrophic lateral sclerosis. Subcommittee on Motor Neuron Diseases/Amyotrophic Lateral Sclerosis of the World Federation of Neurology Research Group on Neuromuscular Diseases and th. J. Neurol. Sci. *124 Suppl*, 96–107.

Brooks, A.R., Harkins, R.N., Wang, P., Qian, H.S., Liu, P., and Rubanyi, G.M. (2004). Transcriptional silencing is associated with extensive methylation of the CMV promoter following adenoviral gene delivery to muscle. J. Gene Med. *6*, 395–404.

Brooks, B.R., Miller, R.G., Swash, M., and Munsat, T.L. (2000). El Escorial revisited: Revised criteria for the diagnosis of amyotrophic lateral sclerosis. Amyotroph. Lateral Scler. *1*, 293–299.

Burberry, A., Suzuki, N., Wang, J.-Y., Moccia, R., Mordes, D.A., Stewart, M.H., Suzuki-Uematsu, S., Ghosh, S., Singh, A., Merkle, F.T., et al. (2016). Loss-of-function mutations in the C9ORF72 mouse ortholog cause fatal autoimmune disease. Sci. Transl. Med. *8*, 347ra93-347ra93.

Carpentier, A.C., Frisch, F., Labbe, S.M., Gagnon, R., De Wal, J., Greentree, S., Petry, H., Twisk, J., Brisson, D., and Gaudet, D. (2012). Effect of alipogene tiparvovec (aav1-lpls447x) on postprandial chylomicron metabolism in Lipoprotein lipase-deficient patients. J. Clin. Endocrinol. Metab. *97*, 1635–1644.

de Carvalho, M., Dengler, R., Eisen, A., England, J.D., Kaji, R., Kimura, J., Mills, K., Mitsumoto, H., Nodera, H., Shefner, J., et al. (2008). Electrodiagnostic criteria for diagnosis of ALS. Clin. Neurophysiol. *119*, 497–503.

Česnik, A.B., Darovic, S., Mihevc, S.P., Štalekar, M., Malnar, M., Motaln, H., Lee, Y.B., Mazej, J., Pohleven, J., Grosch, M., et al. (2019). Nuclear RNA foci from C9ORF72 expansion mutation form paraspeckle-like bodies. J. Cell Sci. *132*.

Challis, R.C., Ravindra Kumar, S., Chan, K.Y., Challis, C., Beadle, K., Jang, M.J., Kim, H.M., Rajendran, P.S., Tompkins, J.D., Shivkumar, K., et al. (2019). Systemic AAV vectors for widespread and targeted gene delivery in rodents. Nat. Protoc. *14*, 379–414.

Chan, K.Y., Jang, M.J., Yoo, B.B., Greenbaum, A., Ravi, N., Wu, W.L., Sánchez-Guardado, L., Lois, C., Mazmanian, S.K., Deverman, B.E., et al. (2017). Engineered AAVs for efficient noninvasive gene delivery to the central and peripheral nervous systems. Nat. Neurosci. *20*, 1172–1179.

Chancellor, A.M., and Warlow, C.P. (1992). Adult onset motor neuron disease: worldwide mortality, incidence and distribution since 1950. J. Neurol. Neurosurg. Psychiatry *55*, 1106–1115.

Charcot, J.-M. (1874). De la Sclérose lateral amyotrophique [French]. Progrès Médical *2*, 325–327, 341–342, 453–455.

Chen, J.S., Dagdas, Y.S., Kleinstiver, B.P., Welch, M.M., Sousa, A.A., Harrington, L.B., Sternberg, S.H., Joung, J.K., Yildiz, A., and Doudna, J.A. (2017). Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. Nature *550*, 407–410.

Chen, Y.Z., Bennett, C.L., Huynh, H.M., Blair, I.P., Puls, I., Irobi, J., Dierick, I., Abel, A., Kennerson, M.L., Rabin, B.A., et al. (2004). DNA/RNA helicase gene mutations in a form of juvenile amyotrophic lateral sclerosis (ALS4). Am. J. Hum. Genet. *74*, 1128–1135.

Cheng, W., Wang, S., Mestre, A.A., Fu, C., Makarem, A., Xian, F., Hayes, L.R., Lopez-Gonzalez, R., Drenner, K., Jiang, J., et al. (2018). C9ORF72 GGGGCC repeat-associated non-AUG translation is upregulated by stress through eIF2α phosphorylation. Nat. Commun. *9*, 51.

Chio, A. (2005). Severely increased risk of amyotrophic lateral sclerosis among Italian professional football players. Brain *128*, 472–476.

Chiò, A., Calvo, A., Dossena, M., Ghiglione, P., Mutani, R., and Mora, G. (2009). ALS in Italian professional soccer players: The risk is still present and could be soccer-specific.

Amyotroph. Lateral Scler. *10*, 205–209.

Chiò, A., Logroscino, G., Traynor, B.J., Collins, J., Simeone, J.C., Goldstein, L.A., and White, L.A. (2013). Global epidemiology of amyotrophic lateral sclerosis: a systematic review of the published literature. Neuroepidemiology *41*, 118–130.

Chiò, A., Battistini, S., and Calvo, A. (2014). Genetic counselling in ALS: facts, uncertainties and clinical suggestions. J Neurol Neurosurg Psychiatry *85*, 478–485.

Cho, S.W., Kim, S., Kim, Y., Kweon, J., Kim, H.S., Bae, S., and Kim, J.S. (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome Res. *24*, 132–141.

Choi, S.Y., Lopez-Gonzalez, R., Krishnan, G., Phillips, H.L., Li, A.N., Seeley, W.W., Yao, W.D., Almeida, S., and Gao, F.B. (2019). C9ORF72-ALS/FTD-associated poly(GR) binds Atp5a1 and compromises mitochondrial function in vivo. Nat. Neurosci. *22*, 851–862.

Choudhury, S.R., Hudry, E., Maguire, C.A., Sena-Esteves, M., Breakefield, X.O., and Grandi, P. (2016). Viral vectors for therapy of neurologic diseases. Neuropharmacology *120*, 63–80.

Chow, C.Y., Landers, J.E., Bergren, S.K., Sapp, P.C., Grant, A.E., Jones, J.M., Everett, L., Lenk, G.M., McKenna-Yasek, D.M., Weisman, L.S., et al. (2009). Deleterious Variants of FIG4, a Phosphoinositide Phosphatase, in Patients with ALS. Am. J. Hum. Genet. *84*, 85–88.

Chow, T.W., Miller, B.L., Hayashi, V.N., and Geschwind, D.H. (1999). Inheritance of frontotemporal dementia. Arch. Neurol. *56*, 817–822.

Cirulli, E.T., Lasseigne, B.N., Petrovski, S., Sapp, P.C., Dion, P.A., Leblond, C.S., Couthouis, J., Lu, Y.F., Wang, Q., Krueger, B.J., et al. (2015). Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. Science (80-.). *347*, 1436–1441.

Ciura, S., Lattante, S., Le Ber, I., Latouche, M., Tostivint, H., Brice, A., and Kabashi, E. (2013). Loss of function of C9orf72 causes motor deficits in a zebrafish model of Amyotrophic Lateral Sclerosis. Ann. Neurol. *74*, 180–187.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems.

Science (80-.). *339*, 819–823.

Cooper-Knock, J., Hewitt, C., Highley, J.R., Brockington, A., Milano, A., Man, S., Martindale, J., Hartley, J., Walsh, T., Gelsthorpe, C., et al. (2012). Clinico-pathological features in amyotrophic lateral sclerosis with expansions in C9ORF72. Brain *135*, 751–764.

Cooper-Knock, J., Walsh, M.J., Higginbottom, A., Robin Highley, J., Dickman, M.J., Edbauer, D., Ince, P.G., Wharton, S.B., Wilson, S.A., Kirby, J., et al. (2014a). Sequestration of multiple RNA recognition motif-containing proteins by C9orf72 repeat expansions. Brain *137*, 2040–2051.

Cooper-Knock, J., Shaw, P.J., and Kirby, J. (2014b). The widening spectrum of C9ORF72-related disease; Genotype/phenotype correlations and potential modifiers of clinical phenotype. Acta Neuropathol. *127*, 333–345.

Cooper-Knock, J., Moll, T., Ramesh, T., Castelli, L., Beer, A., Robins, H., Fox, I., Niedermoser, I., Van Damme, P., Moisse, M., et al. (2019). Mutations in the Glycosyltransferase Domain of GLT8D1 Are Associated with Familial Amyotrophic Lateral Sclerosis. Cell Rep. *26*, 2298-2306.e5.

Coyle-Gilchrist, I.T.S., Dick, K.M., Patterson, K., Rodríquez, P.V., Wehmann, E., Wilcox, A., Lansdall, C.J., Dawson, K.E., Wiggins, J., Mead, S., et al. (2016). Prevalence, characteristics, and survival of frontotemporal lobar degeneration syndromes. Neurology *86*, 1736–1743.

Crockford, C., Newton, J., Lonergan, K., Chiwera, T., Booth, T., Chandran, S., Colville, S., Heverin, M., Mays, I., Pal, S., et al. (2018). ALS-specific cognitive and behavior changes associated with advancing disease stage in ALS. Neurology *91*, E1370–E1380.

Cronin, S., Hardiman, O., and Traynor, B.J. (2007). Ethnic variation in the incidence of ALS: A systematic review. Neurology *68*, 1002–1007.

Cruts, M., Gijselinck, I., Van Der Zee, J., Engelborghs, S., Wils, H., Pirici, D., Rademakers, R., Vandenberghe, R., Dermaut, B., Martin, J.J., et al. (2006). Null mutations in progranulin cause ubiquitin-positive frontotemporal dementia linked to chromosome 17q21. Nature *442*, 920–924.

Cykowski, M.D., Dickson, D.W., Powell, S.Z., Arumanayagam, A.S., Rivera, A.L., and

Appel, S.H. (2019). Dipeptide repeat (DPR) pathology in the skeletal muscle of ALS patients with C9ORF72 repeat expansion. Acta Neuropathol. *138*, 667–670.

Dafinca, R., Scaber, J., Ababneh, N., Lalic, T., Weir, G., Christian, H., Vowles, J., Douglas, A.G.L., Fletcher-Jones, A., Browne, C., et al. (2016). C9orf72 Hexanucleotide Expansions Are Associated with Altered Endoplasmic Reticulum Calcium Homeostasis and Stress Granule Formation in Induced Pluripotent Stem Cell-Derived Neurons from Patients with Amyotrophic Lateral Sclerosis and Frontotemporal Demen. Stem Cells *34*, 2063–2078.

Dafinca, R., Barbagallo, P., Farrimond, L., Candalija, A., Scaber, J., Ababneh, N.A., Sathyaprakash, C., Vowles, J., Cowley, S.A., and Talbot, K. (2020). Impairment of Mitochondrial Calcium Buffering Links Mutations in C9ORF72 and TARDBP in iPS-Derived Motor Neurons from Patients with ALS/FTD. Stem Cell Reports *14*, 892–908.

Deacon, R.M.J. (2013). Measuring the strength of mice. J. Vis. Exp.

Van Deerlin, V.M., Sleiman, P.M.A., Martinez-Lage, M., Chen-Plotkin, A., Wang, L.S., Graff-Radford, N.R., Dickson, D.W., Rademakers, R., Boeve, B.F., Grossman, M., et al. (2010). Common variants at 7p21 are associated with frontotemporal lobar degeneration with TDP-43 inclusions. Nat. Genet. *42*, 234–239.

DeJesus-Hernandez, M., Mackenzie, I.R., Boeve, B.F., Boxer, A.L., Baker, M., Rutherford, N.J., Nicholson, A.M., Finch, N.A., Flynn, H., Adamson, J., et al. (2011). Expanded GGGGCC Hexanucleotide Repeat in Noncoding Region of C9ORF72 Causes Chromosome 9p-Linked FTD and ALS. Neuron *72*, 245–256.

Deng, H.X., Chen, W., Hong, S.T., Boycott, K.M., Gorrie, G.H., Siddique, N., Yang, Y., Fecto, F., Shi, Y., Zhai, H., et al. (2011). Mutations in UBQLN2 cause dominant X-linked juvenile and adult-onset ALS and ALS/dementia. Nature *477*, 211–215.

Deverman, B.E., Pravdo, P.L., Simpson, B.P., Kumar, S.R., Chan, K.Y., Banerjee, A., Wu, W.-L., Yang, B., Huber, N., Pasca, S.P., et al. (2016). Cre-dependent selection yields AAV variants for widespread gene transfer to the adult brain. Nat. Biotechnol. *34*, 204–209.

Dodge, J.C., Haidet, A.M., Yang, W., Passini, M.A., Hester, M., Clarke, J., Roskelley, E.M., Treleaven, C.M., Rizo, L., Martin, H., et al. (2008). Delivery of AAV-IGF-1 to the CNS Extends Survival in ALS Mice Through Modification of Aberrant Glial Cell Activity. Mol. Ther. *16*, 1056–1064.

Dodge, J.C., Treleaven, C.M., Fidler, J.A., Hester, M., Haidet, A., Handy, C., Rao, M., Eagle, A., Matthews, J.C., Taksir, T. V, et al. (2010). AAV4-mediated Expression of IGF-1 and VEGF Within Cellular Components of the Ventricular System Improves Survival Outcome in Familial ALS Mice. Mol. Ther. *18*, 2075–2084.

Dong, F., Xie, K., Chen, Y., Yang, Y., and Mao, Y. (2017). Polycistronic tRNA and CRISPR guide-RNA enables highly efficient multiplexed genome engineering in human cells. Biochem. Biophys. Res. Commun. *482*, 889–895.

Donnelly, C.J., Zhang, P.-W., Pham, J.T., Haeusler, A.R., Mistry, N.A., Vidensky, S., Daley, E.L., Poth, E.M., Hoover, B., Fines, D.M., et al. (2013). RNA Toxicity from the ALS/FTD C9ORF72 Expansion Is Mitigated by Antisense Intervention. Neuron *80*, 415–428.

Dorr, A., Sled, J.G., and Kabani, N. (2007). Three-dimensional cerebral vasculature of the CBA mouse brain: A magnetic resonance imaging and micro computed tomography study. Neuroimage *35*, 1409–1423.

Doudna, J.A., and Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. Science (80-.). *346*.

Duan, W., Guo, M., Yi, L., Liu, Y., Li, Z., Ma, Y., Zhang, G., Liu, Y., Bu, H., Song, X., et al. (2020). The deletion of mutant SOD1 via CRISPR/Cas9/sgRNA prolongs survival in an amyotrophic lateral sclerosis mouse model. Gene Ther. *27*, 157–169.

Duque, S., Joussemet, B., Riviere, C., Marais, T., Dubreil, L., Douar, A.-M., Fyfe, J., Moullier, P., Colle, M.-A., and Barkats, M. (2009). Intravenous Administration of Self-complementary AAV9 Enables Transgene Delivery to Adult Motor Neurons. Mol. Ther. *17*, 1187–1196.

Ebbert, M.T.W., Farrugia, S.L., Sens, J.P., Jansen-West, K., Gendron, T.F., Prudencio, M., McLaughlin, I.J., Bowman, B., Seetin, M., Dejesus-Hernandez, M., et al. (2018). Long-read sequencing across the C9orf72 “GGGGCC” repeat expansion: Implications for clinical use and genetic discovery efforts in human disease. Mol. Neurodegener. *13*, 46.

Elden, A.C., Kim, H.J., Hart, M.P., Chen-Plotkin, A.S., Johnson, B.S., Fang, X., Armakola, M., Geser, F., Greene, R., Lu, M.M., et al. (2010). Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. Nature *466*, 1069–1075.

Elisa, R., Rainero, I., Chiò, A., Rogaeva, E., Galimberti, D., Fenoglio, P., Grinberg, Y., Isaia, G., Calvo, A., Gentile, S., et al. (2012). SQSTM1 mutations in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Neurology *79*, 1556–1562.

van Es, M.A., Hardiman, O., Chio, A., Al-Chalabi, A., Pasterkamp, R.J., Veldink, J.H., and van den Berg, L.H. (2017). Amyotrophic lateral sclerosis. Lancet *390*, 2084–2098.

Escobar-Aguirre, S., Arancibia, D., Escorza, A., Bravo, C., Andrés, M., Zamorano, P., and Martínez, V. (2019). Development of a Bicistronic Vector for the Expression of a CRISPR/Cas9-mCherry System in Fish Cell Lines. Cells *8*, 75.

Farg, M.A., Sundaramoorthy, V., Sultana, J.M., Yang, S., Atkinson, R.A.K., Levina, V., Halloran, M.A., Gleeson, P.A., Blair, I.P., Soo, K.Y., et al. (2014). C9ORF72, implicated in amytrophic lateral sclerosis and frontotemporal dementia, regulates endosomal trafficking. Hum. Mol. Genet. *23*, 3579–3595.

Farg, M.A., Konopka, A., Soo, K.Y., Ito, D., and Atkin, J.D. (2017). The DNA damage response (DDR) is induced by the C9orf72 repeat expansion in amyotrophic lateral sclerosis. Hum. Mol. Genet. *26*, 2882–2896.

Fecto, F., Yan, J., Vemula, S.P., Liu, E., Yang, Y., Chen, W., Zheng, J.G., Shi, Y., Siddique, N., Arrat, H., et al. (2011). SQSTM1 mutations in familial and sporadic amyotrophic lateral sclerosis. Arch. Neurol. *68*, 1440–1446.

Feldman, H., Levy, A.R., Hsiung, G.Y., Peters, K.R., Donald, A., Black, S.E., Bouchard, R.W., Gauthier, S.G., Guzman, D.A., Hogan, D.B., et al. (2003). A Canadian Cohort Study of Cognitive Impairment and Related Dementias (ACCORD): Study methods and baseline results. Neuroepidemiology *22*, 265–274.

Ferrari, R., Kapogiannis, D., Huey, E.D., and Momeni, P. (2011). FTD and ALS: a tale of two diseases. Curr. Alzheimer Res. *8*, 273–294.

Ferrari, R., Hernandez, D.G., Nalls, M.A., Rohrer, J.D., Ramasamy, A., Kwok, J.B.J., Dobson-Stone, C., Brooks William S., B.S., Schofield, P.R., Halliday, G.M., et al. (2014). Frontotemporal dementia and its subtypes: A genome-wide association study. Lancet Neurol. *13*, 686–699.

Figlewicz, D., Krizus, A., Martinoli, M., Meininger, V., Dib, M., Rouleau, G., and Julien, J.-

P. (1994). Variants of the heavy neurofilament subunit are associated with the development of amyotrophic lateral sclerosis. Hum. Mol. Genet. *3*, 1757–1761.

Finkel, R.S., Mercuri, E., Darras, B.T., Connolly, A.M., Kuntz, N.L., Kirschner, J., Chiriboga, C.A., Saito, K., Servais, L., Tizzano, E., et al. (2017). Nusinersen versus Sham Control in Infantile-Onset Spinal Muscular Atrophy. N. Engl. J. Med. *377*, 1723–1732.

Foldvari, M., Chen, D.W., Nafissi, N., Calderon, D., Narsineni, L., and Rafiee, A. (2016). Non-viral gene therapy: Gains and challenges of non-invasive administration methods. J. Control. Release *240*, 165–190.

Forbes, R.B., Colville, S., Parratt, J., and Swingler, R.J. (2007). The incidence of motor nueron disease in Scotland. J. Neurol. *254*, 866–869.

Foust, K.D., Nurre, E., Montgomery, C.L., Hernandez, A., Chan, C.M., and Kaspar, B.K. (2009). Intravascular AAV9 preferentially targets neonatal neurons and adult astrocytes. Nat. Biotechnol. *27*, 59–65.

Foust, K.D., Salazar, D.L., Likhite, S., Ferraiuolo, L., Ditsworth, D., Ilieva, H., Meyer, K., Schmelzer, L., Braun, L., Cleveland, D.W., et al. (2013). Therapeutic AAV9-mediated Suppression of Mutant SOD1 Slows Disease Progression and Extends Survival in Models of Inherited ALS. Mol. Ther. *21*, 2148–2159.

Fratta, P., Poulter, M., Lashley, T., Rohrer, J.D., Polke, J.M., Beck, J., Ryan, N., Hensman, D., Mizielinska, S., Waite, A.J., et al. (2013). Homozygosity for the C9orf72 GGGGCC repeat expansion in frontotemporal dementia. Acta Neuropathol. *126*, 401–409.

Freibaum, B.D., and Taylor, J.P. (2017). The role of dipeptide repeats in C9ORF72-related ALS-FTD. Front. Mol. Neurosci. *10*, 35.

Freibaum, B.D., Lu, Y., Lopez-Gonzalez, R., Kim, N.C., Almeida, S., Lee, K.-H., Badders, N., Valentine, M., Miller, B.L., Wong, P.C., et al. (2015). GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport. Nature *525*, 129–133.

Freischmidt, A., Wieland, T., Richter, B., Ruf, W., Schaeffer, V., Müller, K., Marroquin, N., Nordin, F., Hübers, A., Weydt, P., et al. (2015). Haploinsufficiency of TBK1 causes familial ALS and fronto-temporal dementia. Nat. Neurosci. *18*, 631–636.

Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., and Sander, J.D.

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

Gaj, T., Ojala, D.S., Ekman, F.K., Byrne, L.C., Limsirichai, P., and Schaffer, D. V. (2017). In vivo genome editing improves motor function and extends survival in a mouse model of ALS. Sci. Adv. *3*, eaar3952.

Gallo, V., Bueno-De-Mesquita, H.B., Vermeulen, R., Andersen, P.M., Kyrozis, A., Linseisen, J., Kaaks, R., Allen, N.E., Roddam, A.W., Boshuizen, H.C., et al. (2009). Smoking and risk for amyotrophic lateral sclerosis: Analysis of the EPIC cohort. Ann. Neurol. *65*, 378–385.

Galvin, J.E., Howard, D.H., Denny, S.S., Dickinson, S., and Tatton, N. (2017). The social and economic burden of frontotemporal degeneration. Neurology *89*, 2049–2056.

Gao, Y., and Zhao, Y. (2014). Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing. J. Integr. Plant Biol. *56*, 343–349.

Gaudelli, N.M., Komor, A.C., Rees, H.A., Packer, M.S., Badran, A.H., Bryson, D.I., and Liu, D.R. (2017). Programmable base editing of A T to G C in genomic DNA without DNA cleavage. Nature *551*, 464–471.

Gaudet, D., Méthot, J., Déry, S., Brisson, D., Essiembre, C., Tremblay, G., Tremblay, K., De Wal, J., Twisk, J., Van Den Bulk, N., et al. (2013). 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.

Gauthier, A., Vignola, A., Calvo, A., Cavallo, E., Moglia, C., Sellitti, L., Mutani, R., and Chiò, A. (2007). A longitudinal study on quality of life and depression in ALS patient-caregiver couples. Neurology *68*, 923–926.

Gendron, T.F., and Petrucelli, L. (2018). Disease mechanisms of c9orf72 repeat expansions. Cold Spring Harb. Perspect. Med. *8*.

Gendron, T.F., Bieniek, K.F., Zhang, Y.-J., Jansen-West, K., Ash, P.E.A., Caulfield, T., Daughrity, L., Dunmore, J.H., Castanedes-Casey, M., Chew, J., et al. (2013). Antisense transcripts of the expanded C9ORF72 hexanucleotide repeat form nuclear RNA foci and undergo repeat-associated non-ATG translation in c9FTD/ALS. Acta Neuropathol. *126*, 829–

844.

Gendron, T.F., Belzil, V. V., Zhang, Y.-J., and Petrucelli, L. (2014). Mechanisms of toxicity in C9FTLD/ALS. Acta Neuropathol. *127*, 359–376.

Gendron, T.F., van Blitterswijk, M., Bieniek, K.F., Daughrity, L.M., Jiang, J., Rush, B.K., Pedraza, O., Lucas, J.A., Murray, M.E., Desaro, P., et al. (2015). Cerebellar c9RAN proteins associate with clinical and neuropathological characteristics of C9ORF72 repeat expansion carriers. Acta Neuropathol. *130*, 559–573.

Gendron, T.F., Chew, J., Stankowski, J.N., Hayes, L.R., Zhang, Y.J., Prudencio, M., Carlomagno, Y., Daughrity, L.M., Jansen-West, K., Perkerson, E.A., et al. (2017). Poly(GP) proteins are a useful pharmacodynamic marker for C9ORF72-associated amyotrophic lateral sclerosis. Sci. Transl. Med. *9*, eaai7866.

Geuens, T., Bouhy, D., and Timmerman, V. (2016). The hnRNP family: insights into their role in health and disease. Hum. Genet. *135*, 851–867.

Gholizadeh, S., Tharmalingam, S., MacAldaz, M.E., and Hampson, D.R. (2013). Transduction of the central nervous system after intracerebroventricular injection of adeno-associated viral vectors in neonatal and juvenile mice. Hum. Gene Ther. Methods *24*, 205–213.

Gijselinck, I., Engelborghs, S., Maes, G., Cuijt, I., Peeters, K., Mattheijssens, M., Joris, G., Cras, P., Martin, J.-J., De Deyn, P.P., et al. (2010). Identification of 2 Loci at Chromosomes 9 and 14 in a Multiplex Family With Frontotemporal Lobar Degeneration and Amyotrophic Lateral Sclerosis. Arch. Neurol. *67*, 606–616.

Gijselinck, I., Van Langenhove, T., van der Zee, J., Sleegers, K., Philtjens, S., Kleinberger, G., Janssens, J., Bettens, K., Van Cauwenberghe, C., Pereson, S., et al. (2012). A C9orf72 promoter repeat expansion in a Flanders-Belgian cohort with disorders of the frontotemporal lobar degeneration-amyotrophic lateral sclerosis spectrum: a gene identification study. Lancet. Neurol. *11*, 54–65.

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., et al. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell *154*, 442–451.

Gitler, A.D., and Tsuiji, H. (2016). There has been an awakening: Emerging mechanisms of C9orf72 mutations in FTD/ALS. Brain Res. *1647*, 19–29.

Gladman, M., Dharamshi, C., and Zinman, L. (2014). Economic burden of amyotrophic lateral sclerosis: A Canadian study of out-of-pocket expenses. Amyotroph. Lateral Scler. Front. Degener. *15*, 426–432.

Gleitz, H.F.E., O'Leary, C., Holley, R.J., and Bigger, B.W. (2017). Identification of age-dependent motor and neuropsychological behavioural abnormalities in a mouse model of mucopolysaccharidosis type II. PLoS One *12*.

Goelz, M.F., Mahler, J., Harry, J., Myers, P., Clark, J., Thigpen, J.E., and Forsythe, D.B. (1998). Neuropathologic findings associated with seizures in FVB mice. Lab. Anim. Sci. *48*, 34–37.

Goldman, J.S., Farmer, J.M., Wood, E.M., Johnson, J.K., Boxer, A., Neuhaus, J., Lomen-Hoerth, C., Wilhelmsen, K.C., Lee, V.M.Y., Grossman, M., et al. (2005). Comparison of family histories in FTLD subtypes and related tauopathies. Neurology *65*, 1817–1819.

Goodman, L.D., and Bonini, N.M. (2019). Repeat-associated non-AUG (RAN) translation mechanisms are running into focus for GGGGCC-repeat associated ALS/FTD. Prog. Neurobiol. *183*, 101697.

Gorno-Tempini, M.L., Hillis, A.E., Weintraub, S., Kertesz, A., Mendez, M., Cappa, S.F., Ogar, J.M., Rohrer, J.D., Black, S., Boeve, B.F., et al. (2011). Classification of primary progressive aphasia and its variants. Neurology *76*, 1006–1014.

Goswami, R., Subramanian, G., Silayeva, L., Newkirk, I., Doctor, D., Chawla, K., Chattopadhyay, S., Chandra, D., Chilukuri, N., and Betapudi, V. (2019). Gene therapy leaves a vicious cycle. Front. Oncol. *9*, 297.

Goyenvalle, A., Babbs, A., Wright, J., Wilkins, V., Powell, D., Garcia, L., and Davies, K.E. (2012). Rescue of severely affected dystrophin/utrophin-deficient mice through scAAV-U7snRNA-mediated exon skipping. Hum. Mol. Genet. *21*, 2559–2571.

Goyenvalle, A., Griffith, G., Babbs, A., Andaloussi, S. El, Ezzat, K., Avril, A., Dugovic, B., Chaussenot, R., Ferry, A., Voit, T., et al. (2015). Functional correction in mouse models of muscular dystrophy using exon-skipping tricyclo-DNA oligomers. Nat. Med. *21*, 270–275.

Grabler, M.R., Weyen, U., Juckel, G., Tegenthoff, M., and Mavrogiorgou-Juckel, P. (2018). Death Anxiety and Depression in Amyotrophic Lateral Sclerosis Patients and Their Primary Caregivers. Front. Neurol. *9*, 1035.

Gray, S.J., Foti, S.B., Schwartz, J.W., Bachaboina, L., Taylor-Blake, B., Coleman, J., Ehlers, M.D., Zylka, M.J., McCown, T.J., and Samulski, R.J. (2011). Optimizing Promoters for Recombinant Adeno-Associated Virus-Mediated Gene Expression in the Peripheral and Central Nervous System Using Self-Complementary Vectors. Hum. Gene Ther. *22*, 1143–1153.

Green, K.M., Glineburg, M.R., Kearse, M.G., Flores, B.N., Linsalata, A.E., Fedak, S.J., Goldstrohm, A.C., Barmada, S.J., and Todd, P.K. (2017). RAN translation at C9orf72-associated repeat expansions is selectively enhanced by the integrated stress response. Nat. Commun. *8*, 2005.

Greenway, M.J., Alexander, M.D., Ennis, S., Traynor, B.J., Corr, B., Frost, E., Green, A., and Hardiman, O. (2004). A novel candidate region for ALS on chromosome 14q11.2. Neurology *63*, 1936–1938.

Greenway, M.J., Andersen, P.M., Russ, G., Ennis, S., Cashman, S., Donaghy, C., Patterson, V., Swingler, R., Kieran, D., Prehn, J., et al. (2006). ANG mutations segregate with familial and “sporadic” amyotrophic lateral sclerosis. Nat. Genet. *38*, 411–413.

Gros-Louis, F., Larivière, R., Gowing, G., Laurent, S., Camu, W., Bouchard, J.P., Meininger, V., Rouleau, G.A., and Julien, J.P. (2004). A frameshift deletion in peripherin gene associated with amyotrophic lateral sclerosis. J. Biol. Chem. *279*, 45951–45956.

Grossman, M. (2019). Amyotrophic lateral sclerosis — a multisystem neurodegenerative disorder. Nat. Rev. Neurol. *15*, 5–6.

Gunnarsson, L.-G., Dahlbom, K., and Strandman, E. (1991). Motor neuron disease and dementia reported among 13 members of a single family. Acta Neurol. Scand. *84*, 429–433.

Hadano, S., Hand, C.K., Osuga, H., Yanagisawa, Y., Otomo, A., Devon, R.S., Miyamoto, N., Showguchi-Miyata, J., Okada, Y., Singaraja, R., et al. (2001). A gene encoding a putative GTPase regulator is mutated in familial amyotrophic lateral sclerosis 2. Nat. Genet. *29*, 166–173.

Haeusler, A.R., Donnelly, C.J., Periz, G., Simko, E.A.J., Shaw, P.G., Kim, M.-S., Maragakis, N.J., Troncoso, J.C., Pandey, A., Sattler, R., et al. (2014). C9orf72 nucleotide repeat structures initiate molecular cascades of disease. Nature *507*, 195–200.

Haeusler, A.R., Donnelly, C.J., and Rothstein, J.D. (2016). The expanding biology of the C9orf72 nucleotide repeat expansion in neurodegenerative disease. Nat. Rev. Neurosci. *17*, 383–395.

Haeussler, M., Schönig, K., Eckert, H., Eschstruth, A., Mianné, J., Renaud, J.B., Schneider-Maunoury, S., Shkumatava, A., Teboul, L., Kent, J., et al. (2016). Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. Genome Biol. *17*, 148.

Haley, R.W. (2003). Excess incidence of ALS in young Gulf War veterans. Neurology *61*, 750–756.

Hand, C.K., Khoris, J., Salachas, F., Gros-Louis, F., Simões Lopes, A.A., Mayeux-Portas, V., Brown, R.H., Meininger, V., Camu, W., and Rouleau, G.A. (2002). A novel locus for familial amyotrophic lateral sclerosis, on chromosome 18q. Am. J. Hum. Genet. *70*, 251–256.

Hantash, F.M., Goos, D.G., Tsao, D., Quan, F., Buller-Burckle, A., Peng, M., Jarvis, M., Sun, W., and Strom, C.M. (2010). Qualitative assessment of fmr1 (CGG)n triplet repeat status in normal, intermediate, premutation, full mutation, and mosaic carriers in both sexes: Implications for fragile x syndrome carrier and newborn screening. Genet. Med. *12*, 162–173.

Hardiman, O., van den Berg, L.H., and Kiernan, M.C. (2011). Clinical diagnosis and management of amyotrophic lateral sclerosis. Nat. Rev. Neurol. *7*, 639–649.

Hardiman, O., Al-Chalabi, A., Chio, A., Corr, E.M., Logroscino, G., Robberecht, W., Shaw, P.J., Simmons, Z., and Van Den Berg, L.H. (2017). Amyotrophic lateral sclerosis. Nat. Rev. Dis. Prim. *3*, 1–19.

Harvey, R.J., Skelton-Robinson, M., and Rossor, M.N. (2003). The prevalence and causes of dementia in people under the age of 65 years. J. Neurol. Neurosurg. Psychiatry *74*, 1206–1209.

Hautbergue, G.M., Castelli, L.M., Ferraiuolo, L., Sanchez-Martinez, A., Cooper-Knock, J., Higginbottom, A., Lin, Y.-H., Bauer, C.S., Dodd, J.E., Myszczynska, M.A., et al. (2017).

SRSF1-dependent nuclear export inhibition of C9ORF72 repeat transcripts prevents neurodegeneration and associated motor deficits. Nat. Commun. *8*, 16063.

Hayes, L.R., and Rothstein, J.D. (2016). C9ORF72-ALS/FTD: Transgenic Mice Make a Come-BAC. Neuron *90*, 427–431.

Heckl, D., Kowalczyk, M.S., Yudovich, D., Belizaire, R., Puram, R. V., McConkey, M.E., Thielke, A., Aster, J.C., Regev, A., and Ebert, B.L. (2014). Generation of mouse models of myeloid malignancy with combinatorial genetic lesions using CRISPR-Cas9 genome editing. Nat. Biotechnol. *32*, 941–946.

Henriques, A., Pitzer, C., Dittgen, T., Klugmann, M., Dupuis, L., and Schneider, A. (2011). CNS-targeted Viral Delivery of G-CSF in an Animal Model for ALS: Improved Efficacy and Preservation of the Neuromuscular Unit. Mol. Ther. *19*, 284–292.

Hentati, A., Bejaoui, K., Pericak-Vance, M.A., Hentati, F., Speer, M.C., Hung, W.Y., Figlewicz, D.A., Haines, J., Rimmler, J., Hamida, C. Ben, et al. (1994). Linkage of recessive familial amyotrophic lateral sclerosis to chromosome 2q33–q35. Nat. Genet. *7*, 425–428.

Herranz-Martin, S., Chandran, J., Lewis, K., Mulcahy, P., Higginbottom, A., Walker, C., Valenzuela, I.M.-P.Y., Jones, R.A., Coldicott, I., Iannitti, T., et al. (2017). Viral delivery of C9orf72 hexanucleotide repeat expansions in mice leads to repeat-length-dependent neuropathology and behavioural deficits. Dis. Model. Mech. *10*, 859–868.

Heutink, P., Jansen, I.E., and Lynes, E.M. (2014). C9orf72; abnormal RNA expression is the key. Exp. Neurol. *262*, 102–110.

Hille, F., and Charpentier, E. (2016). CRISPR-Cas: biology, mechanisms and relevance. Philos. Trans. R. Soc. London B Biol. Sci. *371*.

Hinderer, C., Katz, N., Buza, E.L., Dyer, C., Goode, T., Bell, P., Richman, L.K., and Wilson, J.M. (2018). Severe Toxicity in Nonhuman Primates and Piglets Following High-Dose Intravenous Administration of an Adeno-Associated Virus Vector Expressing Human SMN. Hum. Gene Ther. *29*, 285–298.

Hodges, J.R., Davies, R., Xuereb, J., Kril, J., and Halliday, G. (2003). Survival in frontotemporal dementia. Neurology *61*, 349–354.

Hoffmann, D., Schott, J.W., Geis, F.K., Lange, L., Müller, F.J., Lenz, D., Zychlinski, D.,

Steinemann, D., Morgan, M., Moritz, T., et al. (2017). Detailed comparison of retroviral vectors and promoter configurations for stable and high transgene expression in human induced pluripotent stem cells. Gene Ther. *24*, 298–307.

Holehonnur, R., Lella, S.K., Ho, A., Luong, J.A., and Ploski, J.E. (2015). The production of viral vectors designed to express large and difficult to express transgenes within neurons. Mol. Brain *8*, 12.

Holm, I.E., Englund, E., Mackenzie, I.R.A., Johannsen, P., and Isaacs, A.M. (2007). A reassessment of the neuropathology of frontotemporal dementia linked to chromosome 3. J. Neuropathol. Exp. Neurol. *66*, 884–891.

Holm, I.E., Isaacs, A.M., and MacKenzie, I.R.A. (2009). Absence of FUS-immunoreactive pathology in frontotemporal dementia linked to chromosome 3 (FTD-3) caused by mutation in the CHMP2B gene. Acta Neuropathol. *118*, 719–720.

Hordeaux, J., Wang, Q., Katz, N., Buza, E.L., Bell, P., and Wilson, J.M. (2018). The Neurotropic Properties of AAV-PHP.B Are Limited to C57BL/6J Mice. Mol. Ther. *26*, 664–668.

Hordeaux, J., Yuan, Y., Clark, P.M., Wang, Q., Martino, R.A., Sims, J.J., Bell, P., Raymond, A., Stanford, W.L., and Wilson, J.M. (2019). The GPI-Linked Protein LY6A Drives AAV-PHP.B Transport across the Blood-Brain Barrier. Mol. Ther. *27*, 912–921.

Horner, R.D., Kamins, K.G., Feussner, J.R., Grambow, S.C., Hoff-Lindquist, J., Harati, Y., Mitsumoto, H., Pascuzzi, R., Spencer, P.S., Tim, R., et al. (2003). Occurrence of amyotrophic lateral sclerosis among Gulf War veterans. Neurology *61*, 742–749.

Hosler, B.A., Siddique, T., Sapp, P.C., Sailor, W., Huang, M.C., Hossain, A., Daube, J.R., Nance, M., Fan, C., Kaplan, J., et al. (2000). Linkage of familial amyotrophic lateral sclerosis with frontotemporal dementia to chromosome 9q21-q22. JAMA *284*, 1664–1669.

Hoy, S.M. (2019). Onasemnogene Abeparvovec: First Global Approval. Drugs *79*, 1255–1262.

Hsu, C.C., Li, H.P., Hung, Y.H., Leu, Y.W., Wu, W.H., Wang, F.S., Lee, K. Der, Chang, P.J., Wu, C.S., Lu, Y.J., et al. (2010). Targeted methylation of CMV and E1A viral promoters. Biochem. Biophys. Res. Commun. *402*, 228–234.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nat. Biotechnol. *31*, 827–832.

Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and Applications of CRISPR-Cas9 for Genome Engineering. Cell *157*, 1262–1278.

Hu, C., Busuttil, R.W., and Lipshutz, G.S. (2010). RH10 provides superior transgene expression in mice when compared with natural AAV serotypes for neonatal gene therapy. J. Gene Med. *12*, 766–778.

Huang, Q., Chan, K.Y., Tobey, I.G., Chan, Y.A., Poterba, T., Boutros, C.L., Balazs, A.B., Daneman, R., Bloom, J.M., Seed, C., et al. (2019). Delivering genes across the blood-brain barrier: LY6A, a novel cellular receptor for AAV-PHP.B capsids. PLoS One *14*.

Hui, L., Zhao, M., He, J., Hu, Y., Huo, Y., Hao, H., Hao, Y., Zhu, W., Wang, Y., Xu, M., et al. (2019). A simple and reliable method for creating PCR-detectable mutants in Arabidopsis with the polycistronic tRNA–gRNA CRISPR/Cas9 system. Acta Physiol. Plant. *41*, 170.

Hutton, M., Lendon, C.L., Rizzu, P., Baker, M., Froelich, S., Houlden, H.H., Pickering-Brown, S., Chakraverty, S., Isaacs, A., Grover, A., et al. (1998). Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17. Nature *393*, 702–704.

Hvidsten, L., Engedal, K., Selbæk, G., Wyller, T.B., Benth, J.Š., and Kersten, H. (2019). Quality of life of family carers of persons with young-onset dementia: A Nordic two-year observational multicenter study. PLoS One *14*.

Iannitti, T., Scarrott, J.M., Likhite, S., Coldicott, I.R.P., Lewis, K.E., Heath, P.R., Higginbottom, A., Myszczynska, M.A., Milo, M., Hautbergue, G.M., et al. (2018). Translating SOD1 Gene Silencing toward the Clinic: A Highly Efficacious, Off-Target-free, and Biomarker-Supported Strategy for fALS. Mol. Ther. Nucleic Acids *12*, 75–88.

Ikejima, C., Yasuno, F., Mizukami, K., Sasaki, M., Tanimukai, S., and Asada, T. (2009). Prevalence and causes of early-onset dementia in japan: A population-based study. Stroke *40*, 2709–2714.

Inagaki, K., Fuess, S., Storm, T.A., Gibson, G.A., Mctiernan, C.F., Kay, M.A., and Nakai, H. (2006). Robust systemic transduction with AAV9 vectors in mice: efficient global cardiac

gene transfer superior to that of AAV8. Mol. Ther. *14*, 45–53.

Ingusci, S., Verlengia, G., Soukupova, M., Zucchini, S., and Simonato, M. (2019). Gene Therapy Tools for Brain Diseases. Front. Pharmacol. *10*, 724.

Ishikawa, K., Weber, T., and Hajjar, R.J. (2018). Human cardiac gene therapy. Circ. Res. *123*, 601–613.

Jaiswal, M.K. (2019). Riluzole and edaravone: A tale of two amyotrophic lateral sclerosis drugs. Med. Res. Rev. *39*, 733–748.

Jayant, R.D., Sosa, D., Kaushik, A., Atluri, V., Vashist, A., Tomitaka, A., and Nair, M. (2016). Current status of non-viral gene therapy for CNS disorders. Expert Opin. Drug Deliv. *13*, 1433–1445.

Jiang, J., Zhu, Q., Gendron, T.F., Saberi, S., McAlonis-Downes, M., Seelman, A., Stauffer, J.E., Jafar-nejad, P., Drenner, K., Schulte, D., et al. (2016). Gain of Toxicity from ALS/FTD-Linked Repeat Expansions in C9ORF72 Is Alleviated by Antisense Oligonucleotides Targeting GGGGCC-Containing RNAs. Neuron *90*, 535–550.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science (80-.). *337*, 816–821.

Johnson, F.O., and Atchison, W.D. (2009). The role of environmental mercury, lead and pesticide exposure in development of amyotrophic lateral sclerosis. Neurotoxicology *30*, 761–765.

Johnson, J.O., Mandrioli, J., Benatar, M., Abramzon, Y., Van Deerlin, V.M., Trojanowski, J.Q., Gibbs, J.R., Brunetti, M., Gronka, S., Wuu, J., et al. (2010). Exome Sequencing Reveals VCP Mutations as a Cause of Familial ALS. Neuron *68*, 857–864.

Johnson, J.O., Pioro, E.P., Boehringer, A., Chia, R., Feit, H., Renton, A.E., Pliner, H.A., Abramzon, Y., Marangi, G., Winborn, B.J., et al. (2014). Mutations in the Matrin 3 gene cause familial amyotrophic lateral sclerosis. Nat. Neurosci. *17*, 664–666.

Josephs, K.A., Whitwell, J.L., Parisi, J.E., Petersen, R.C., Boeve, B.F., Jack, C.R., and Dickson, D.W. (2010). Caudate atrophy on MRI is a characteristic feature of FTLD-FUS. Eur. J. Neurol. *17*, 969–975.

Josephs, K.A., Hodges, J.R., Snowden, J.S., MacKenzie, I.R., Neumann, M., Mann, D.M., and Dickson, D.W. (2011). Neuropathological background of phenotypical variability in frontotemporal dementia. Acta Neuropathol. *122*, 137–153.

Joyce, N.C., and Carter, G.T. (2013). Electrodiagnosis in amyotrophic lateral sclerosis. PM R *5*, S89–S95.

JR, D. (2000). Electrodiagnostic Studies in Amyotrophic Lateral Sclerosis and Other Motor Neuron Disorders. Muscle Nerve *23*, 1488–1502.

Jun, K.Y., Park, J., Oh, K.W., Kim, E.M., Bae, J.S., Kim, I., and Kim, S.H. (2019). Epidemiology of ALS in Korea using nationwide big data. J. Neurol. Neurosurg. Psychiatry *90*, 395–403.

Kabashi, E., Valdmanis, P.N., Dion, P., Spiegelman, D., McConkey, B.J., Velde, C. Vande, Bouchard, J.P., Lacomblez, L., Pochigaeva, K., Salachas, F., et al. (2008). TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. Nat. Genet. *40*, 572–574.

Kadoyama, K., Funakoshi, H., Ohya, W., and Nakamura, T. (2007). Hepatocyte growth factor (HGF) attenuates gliosis and motoneuronal degeneration in the brainstem motor nuclei of a transgenic mouse model of ALS. Neurosci. Res. *59*, 446–456.

Kamel, F., Umbach, D.M., Bedlack, R.S., Richards, M., Watson, M., Alavanja, M.C.R., Blair, A., Hoppin, J.A., Schmidt, S., and Sandler, D.P. (2012). Pesticide exposure and amyotrophic lateral sclerosis. Neurotoxicology *33*, 457–462.

Karimian, A., Azizian, K., Parsian, H., Rafieian, S., Shafiei-Irannejad, V., Kheyrollah, M., Yousefi, M., Majidinia, M., and Yousefi, B. (2019). CRISPR/Cas9 technology as a potent molecular tool for gene therapy. J. Cell. Physiol. *234*, 12267–12277.

Karra, D., and Dahm, R. (2010). Transfection techniques for neuronal cells. J. Neurosci. *30*, 6171–6177.

Kasarskis, E.J., Lindquist, J.H., Coffman, C.J., Grambow, S.C., Feussner, J.R., Allen, K.D., Oddone, E.Z., Kamins, K.A., Horner, R.D., and Als Gulf War Clinical Review Team (2009). Clinical aspects of ALS in Gulf War Veterans. Amyotroph. Lateral Scler. *10*, 35–41.

Kaspar, B.K., Lladó, J., Sherkat, N., Rothstein, J.D., and Gage, F.H. (2003). Retrograde Viral

Delivery of IGF-1 Prolongs Survival in a Mouse ALS Model. Science (80-.). *301*, 839–842.

Kato, S., Shimoda, M., Watanabe, Y., Nakashima, K., Takahashi, K., and Ohama, E. (1996). Familial amyotrophic lateral sclerosis with a two base pair deletion in superoxide dismutase 1: gene multisystem degeneration with intracytoplasmic hyaline inclusions in astrocytes. J. Neuropathol. Exp. Neurol. *55*, 1089–1101.

Keeler, A.M., Zieger, M., Semple, C., Pucci, L., Veinbachs, A., Brown, R.H., Mueller, C., and ElMallah, M.K. (2020). Intralingual and Intrapleural AAV Gene Therapy Prolongs Survival in a SOD1 ALS Mouse Model. Mol. Ther. - Methods Clin. Dev. *17*, 246–257.

Kells, A.P., Hadaczek, P., Yin, D., Bringas, J., Varenika, V., Forsayeth, J., and Bankiewicz, K.S. (2009). Efficient gene therapy-based method for the delivery of therapeutics to primate cortex. Proc. Natl. Acad. Sci. U. S. A. *106*, 2407–2411.

Kenna, K.P., Van Doormaal, P.T.C., Dekker, A.M., Ticozzi, N., Kenna, B.J., Diekstra, F.P., Van Rheenen, W., Van Eijk, K.R., Jones, A.R., Keagle, P., et al. (2016). NEK1 variants confer susceptibility to amyotrophic lateral sclerosis. Nat. Genet. *48*, 1037–1042.

Kiernan, M.C., Vucic, S., Cheah, B.C., Turner, M.R., Eisen, A., Hardiman, O., Burrell, J.R., and Zoing, M.C. (2011). Amyotrophic lateral sclerosis. Lancet *377*, 942–955.

Kim, D., Bae, S., Park, J., Kim, E., Kim, S., Yu, H.R., Hwang, J., Kim, J. Il, and Kim, J.S. (2015). Digenome-seq: Genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. Nat. Methods *12*, 237–243.

Kim, D.W., Uetsuki, T., Kaziro, Y., Yamaguchi, N., and Sugano, S. (1990). Use of the human elongation factor 1α promoter as a versatile and efficient expression system. Gene *91*, 217–223.

Kim, E., Koo, T., Park, S.W., Kim, D., Kim, K., Cho, H.Y., Song, D.W., Lee, K.J., Jung, M.H., Kim, S., et al. (2017). In vivo genome editing with a small Cas9 orthologue derived from Campylobacter jejuni. Nat. Commun. *8*, 14500.

Kim, H.J., Kim, N.C., Wang, Y.D., Scarborough, E.A., Moore, J., Diaz, Z., MacLea, K.S., Freibaum, B., Li, S., Molliex, A., et al. (2013). Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. Nature *495*, 467–473.

Kinsley, L., and Siddique, T. (1993). Amyotrophic Lateral Sclerosis Overview (University of Washington, Seattle).

Klein, R.L., Meyer, E.M., Peel, A.L., Zolotukhin, S., Meyers, C., Muzyczka, N., and King, M.A. (1998). Neuron-specific transduction in the rat septohippocampal or nigrostriatal pathway by recombinant adeno-associated virus vectors. Exp. Neurol. *150*, 183–194.

Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z., and Joung, J.K. (2016). High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. Nature *529*, 490–495.

Knopman, D.S., and Roberts, R.O. (2011). Estimating the number of persons with frontotemporal lobar degeneration in the US population. J. Mol. Neurosci. *45*, 330–335.

Kohnken, R.A., and Schwahn, D.J. (2016). Lack of Chronic Histologic Lesions Supportive of Sublethal Spontaneous Seizures in FVB/N Mice. Comp. Med. *66*, 105–111.

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

Kondziella, W. (1964). A new method for the measurement of muscle relaxation in white mice. Arch. Int. Pharmacodyn. Ther. *152*, 277–284.

Koppers, M., Blokhuis, A.M., Westeneng, H.-J., Terpstra, M.L., Zundel, C.A.C., Vieira de Sá, R., Schellevis, R.D., Waite, A.J., Blake, D.J., Veldink, J.H., et al. (2015). C9orf72 ablation in mice does not cause motor neuron degeneration or motor deficits. Ann. Neurol. *78*, 426–438.

Kotterman, M.A., Chalberg, T.W., and Schaffer, D. V. (2015). Viral Vectors for Gene Therapy: Translational and Clinical Outlook. Annu. Rev. Biomed. Eng. *17*, 63–89.

Kovacs, G.G., Murrell, J.R., Horvath, S., Haraszti, L., Majtenyi, K., Molnar, M.J., Budka, H., Ghetti, B., and Spina, S. (2009). TARDBP variation associated with frontotemporal dementia, supranuclear gaze palsy, and chorea. Mov. Disord. *24*, 1842–1847.

Koval, E.D., Shaner, C., Zhang, P., du Maine, X., Fischer, K., Tay, J., Chau, B.N., Wu, G.F., and Miller, T.M. (2013). Method for widespread microRNA-155 inhibition prolongs survival in ALS-model mice. Hum. Mol. Genet. *22*, 4127–4135.

Krakora, D., Mulcrone, P., Meyer, M., Lewis, C., Bernau, K., Gowing, G., Zimprich, C., Aebischer, P., Svendsen, C.N., and Suzuki, M. (2013). Synergistic Effects of GDNF and VEGF on Lifespan and Disease Progression in a Familial ALS Rat Model. Mol. Ther. *21*, 1602–1610.

Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E., and Cech, T.R. (1982). Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. Cell *31*, 147–157.

Kumar, N., Stanford, W., de Solis, C., Aradhana, Abraham, N.D., Dao, T.M.J., Thaseen, S., Sairavi, A., Gonzalez, C.U., and Ploski, J.E. (2018). The development of an AAV-based crispr sacas9 genome editing system that can be delivered to neurons in vivo and regulated via doxycycline and cre-recombinase. Front. Mol. Neurosci. *11*, 413.

Kwiatkowski, T.J., Bosco, D.A., LeClerc, A.L., Tamrazian, E., Vanderburg, C.R., Russ, C., Davis, A., Gilchrist, J., Kasarskis, E.J., Munsat, T., et al. (2009). Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. Science (80-.). *323*, 1205–1208.

Kwon, I., Xiang, S., Kato, M., Wu, L., Theodoropoulos, P., Wang, T., Kim, J., Yun, J., Xie, Y., and McKnight, S.L. (2014). Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. Science (80-.). *345*, 1139–1145.

Lagier-Tourenne, C., Baughn, M., Rigo, F., Sun, S., Liu, P., Li, H.-R., Jiang, J., Watt, A.T., Chun, S., Katz, M., et al. (2013). Targeted degradation of sense and antisense C9orf72 RNA foci as therapy for ALS and frontotemporal degeneration. Proc. Natl. Acad. Sci. U. S. A. *110*, E4530–E4539.

Lambert, M.A., Bickel, H., Prince, M., Fratiglioni, L., Von Strauss, E., Frydecka, D., Kiejna, A., Georges, J., and Reynish, E.L. (2014). Estimating the burden of early onset dementia; systematic review of disease prevalence. Eur. J. Neurol. *21*, 563–569.

Van Langenhove, T., Van Der Zee, J., Sleegers, K., Engelborghs, S., Vandenberghe, R., Gijselinck, I., Van Den Broeck, M., Mattheijssens, M., Peeters, K., De Deyn, P.P., et al. (2010). Genetic contribution of FUS to frontotemporal lobar degeneration. Neurology *74*, 366–371.

Lee, E.B., Porta, S., Michael Baer, G., Xu, Y., Suh, E., Kwong, L.K., Elman, L., Grossman,

M., Lee, V.M.-Y., Irwin, D.J., et al. (2017). Expansion of the classification of FTLD-TDP: distinct pathology associated with rapidly progressive frontotemporal degeneration. Acta Neuropathol. *134*, 65–78.

Lee, K.-H., Zhang, P., Kim, H.J., Mitrea, D.M., Sarkar, M., Freibaum, B.D., Cika, J., Coughlin, M., Messing, J., Molliex, A., et al. (2016). C9orf72 Dipeptide Repeats Impair the Assembly, Dynamics, and Function of Membrane-Less Organelles. Cell *167*, 774-788.e17.

Lehmer, C., Oeckl, P., Weishaupt, J.H., Volk, A.E., Diehl-Schmid, J., Schroeter, M.L., Lauer, M., Kornhuber, J., Levin, J., Fassbender, K., et al. (2017). Poly- GP in cerebrospinal fluid links C9orf72 -associated dipeptide repeat expression to the asymptomatic phase of ALS / FTD . EMBO Mol. Med. *9*, 859–868.

Leighton, D.J., Newton, J., Stephenson, L.J., Colville, S., Davenport, R., Gorrie, G., Morrison, I., Swingler, R., Chandran, S., and Pal, S. (2019). Changing epidemiology of motor neurone disease in Scotland. J. Neurol. *266*, 817–825.

Levine, T.P., Daniels, R.D., Gatta, A.T., Wong, L.H., and Hayes, M.J. (2013). The product of C9orf72, a gene strongly implicated in neurodegeneration, is structurally related to DENN Rab-GEFs. Bioinformatics *29*, 499–503.

Levitt, N., Briggs, D., Gil, A., and Proudfoot, N.J. (1989). Definition of an efficient synthetic poly(A) site. Genes Dev. *3*, 1019–1025.

Li, Y., and Jin, P. (2012). RNA-mediated neurodegeneration in fragile X-associated tremor/ataxia syndrome. Brain Res. *1462*, 112–117.

Lim, C.K.W., Gapinske, M., Brooks, A.K., Woods, W.S., Powell, J.E., Zeballos C., M.A., Winter, J., Perez-Pinera, P., and Gaj, T. (2020). Treatment of a Mouse Model of ALS by In Vivo Base Editing. Mol. Ther. *28*, 1177–1189.

Lin, Y., Mori, E., Kato, M., Xiang, S., Wu, L., Kwon, I., and McKnight, S.L. (2016). Toxic PR Poly-Dipeptides Encoded by the C9orf72 Repeat Expansion Target LC Domain Polymers. Cell *167*, 789-802.e12.

Liu, E.Y., Russ, J., Wu, K., Neal, D., Suh, E., McNally, A.G., Irwin, D.J., Van Deerlin, V.M., and Lee, E.B. (2014). C9orf72 hypermethylation protects against repeat expansion-associated pathology in ALS/FTD. Acta Neuropathol. *128*, 525–541.

Liu, J., Hu, J., Ludlow, A.T., Pham, J.T., Shay, J.W., Rothstein, J.D., and Corey, D.R. (2017). c9orf72 Disease-Related Foci Are Each Composed of One Mutant Expanded Repeat RNA. Cell Chem. Biol. *24*, 141–148.

Liu, Y., Pattamatta, A., Zu, T., Reid, T., Bardhi, O., Borchelt, D.R., Yachnis, A.T., and Ranum, L.P.W. (2016). C9orf72 BAC Mouse Model with Motor Deficits and Neurodegenerative Features of ALS/FTD. Neuron *90*, 521–534.

Locatelli, F., Corti, S., Papadimitriou, D., Fortunato, F., Del Bo, R., Donadoni, C., Nizzardo, M., Nardini, M., Salani, S., Ghezzi, S., et al. (2007). Fas small interfering RNA reduces motoneuron death in amyotrophic lateral sclerosis mice. Ann. Neurol. *62*, 81–92.

Logroscino, G., Beghi, E., Zoccolella, S., Palagano, R., Fraddosio, A., Simone, I.L., Lamberti, P., Lepore, V., Serlenga, L., and SLAP Registry, S. (2005). Incidence of amyotrophic lateral sclerosis in southern Italy: a population based study. J. Neurol. Neurosurg. Psychiatry *76*, 1094–1098.

Logroscino, G., Traynor, B.J., Hardiman, O., Chio, A., Mitchell, D., Swingler, R.J., Millul, A., Benn, E., and Beghi, E. (2010). Incidence of amyotrophic lateral sclerosis in Europe. J. Neurol. Neurosurg. Psychiatry *81*, 385–390.

Lomen-Hoerth, C., Murphy, J., Langmore, S., Kramer, J.H., Olney, R.K., and Miller, B. (2003). Are amyotrophic lateral sclerosis patients cognitively normal? Neurology *60*, 1094–1097.

Long, C., McAnally, J.R., Shelton, J.M., Mireault, A.A., Bassel-Duby, R., and Olson, E.N. (2014). Prevention of muscular dystrophy in mice by CRISPR/Cas9-mediated editing of germline DNA. Science (80-.). *345*, 1184–1188.

Long, C., Amoasii, L., Mireault, A.A., McAnally, J.R., Li, H., Sanchez-Ortiz, E., Bhattacharyya, S., Shelton, J.M., Bassel-Duby, R., and Olson, E.N. (2016). Postnatal genome editing partially restores dystrophin expression in a mouse model of muscular dystrophy. Science (80-.). *351*, 400–403.

Longinetti, E., and Fang, F. (2019). Epidemiology of amyotrophic lateral sclerosis: An update of recent literature. Curr. Opin. Neurol. *32*, 771–776.

Longinetti, E., Regodón Wallin, A., Samuelsson, K., Press, R., Zachau, A., Ronnevi, L.O.,

Kierkegaard, M., Andersen, P.M., Hillert, J., Fang, F., et al. (2018). The Swedish motor neuron disease quality registry. Amyotroph. Lateral Scler. Front. Degener. *19*, 528–537.

Lopez-Gonzalez, R., Lu, Y., Gendron, T.F., Miller, B.L., Almeida, S., Gao, F.-B., Karydas, A., Tran, H., Yang, D., and Petrucelli, L. (2016). Poly(GR) in C9ORF72-Related ALS/FTD Compromises Mitochondrial Function and Increases Oxidative Stress and DNA Damage in iPSC-Derived Motor Neurons Correspondence Poly(GR) in C9ORF72-Related ALS/FTD Compromises Mitochondrial Function and Increases Oxidative Stress and DNA Damage in iPSC-Derived Motor Neurons. Neuron *92*, 383–391.

Loughner, C.L., Bruford, E.A., McAndrews, M.S., Delp, E.E., Swamynathan, S., and Swamynathan, S.K. (2016). Organization, evolution and functions of the human and mouse Ly6/uPAR family genes. Hum. Genomics *10*, 10.

Lukashchuk, V., Lewis, K.E., Coldicott, I., Grierson, A.J., and Azzouz, M. (2016). AAV9-mediated central nervous system–targeted gene delivery via cisterna magna route in mice. Mol. Ther. - Methods Clin. Dev. *3*, 15055.

Lusis, A.J., Yu, J., and Wang, S.S. (2007). The problem of passenger genes in transgenic mice. Arterioscler. Thromb. Vasc. Biol. *27*, 2100–2103.

Luty, A.A., Kwok, J.B.J., Dobson-Stone, C., Loy, C.T., Coupland, K.G., Karlström, H., Sobow, T., Tchorzewska, J., Maruszak, A., Barcikowska, M., et al. (2010). Sigma nonopioid intracellular receptor 1 mutations cause frontotemporal lobar degeneration-motor neuron disease. Ann. Neurol. *68*, 639–649.

Mackenzie, I.R.A., Neumann, M., Bigio, E.H., Cairns, N.J., Alafuzoff, I., Kril, J., Kovacs, G.G., Ghetti, B., Halliday, G., Holm, I.E., et al. (2009). Nomenclature for neuropathologic subtypes of frontotemporal lobar degeneration: Consensus recommendations. Acta Neuropathol. *117*, 15–18.

Mackenzie, I.R.A., Neumann, M., Baborie, A., Sampathu, D.M., Du Plessis, D., Jaros, E., Perry, R.H., Trojanowski, J.Q., Mann, D.M.A., and Lee, V.M.Y. (2011). A harmonized classification system for FTLD-TDP pathology. Acta Neuropathol. *122*, 111–113.

Mackenzie, I.R.A., Frick, P., Grässer, F.A., Gendron, T.F., Petrucelli, L., Cashman, N.R., Edbauer, D., Kremmer, E., Prudlo, J., Troost, D., et al. (2015). Quantitative analysis and clinico-pathological correlations of different dipeptide repeat protein pathologies in

C9ORF72 mutation carriers. Acta Neuropathol. *130*, 845–861.

MacKenzie, I.R.A., Neumann, M., Bigio, E.H., Cairns, N.J., Alafuzoff, I., Kril, J., Kovacs, G.G., Ghetti, B., Halliday, G., Holm, I.E., et al. (2010). Nomenclature and nosology for neuropathologic subtypes of frontotemporal lobar degeneration: An update. Acta Neuropathol. *119*, 1–4.

MacKenzie, I.R.A., Munoz, D.G., Kusaka, H., Yokota, O., Ishihara, K., Roeber, S., Kretzschmar, H.A., Cairns, N.J., and Neumann, M. (2011). Distinct pathological subtypes of FTLD-FUS. Acta Neuropathol. *121*, 207–218.

MacLaren, R.E., Groppe, M., Barnard, A.R., Cottriall, C.L., Tolmachova, T., Seymour, L., Reed Clark, K., During, M.J., Cremers, F.P.M., Black, G.C.M., et al. (2014). Retinal gene therapy in patients with choroideremia: Initial fi ndings from a phase 1/2 clinical trial. Lancet *383*, 1129–1137.

Maeder, M.L., and Gersbach, C.A. (2016). Genome-editing Technologies for Gene and Cell Therapy. Mol. Ther. *24*, 430–446.

Maguire, C.A., Ramirez, S.H., Merkel, S.F., Sena-Esteves, M., and Breakefield, X.O. (2014). Gene Therapy for the Nervous System: Challenges and New Strategies. Neurotherapeutics *11*, 817–839.

Mahler, J.F., Stokes, W., Mann, P.C., Takaoka, M., and Maronpot, R.R. (1996). Spontaneous lesions in aging FVB/N mice. Toxicol. Pathol. *24*, 710–716.

Majoor-Krakauer, D., Ottman, R., Johnson, W.G., and Rowland, L.P. (1994). Familial aggregation of amyotrophic lateral sclerosis, dementia, and Parkinson's disease: Evidence of shared genetic susceptibility. Neurology *44*, 1872–1877.

Majounie, E., Renton, A.E., Mok, K., Dopper, E.G.P., Waite, A., Rollinson, S., Chiò, A., Restagno, G., Nicolaou, N., Simon-Sanchez, J., et al. (2012). Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a cross-sectional study. Lancet. Neurol. *11*, 323–330.

Malek, A.M., Barchowsky, A., Bowser, R., Youk, A., and Talbott, E.O. (2012). Pesticide exposure as a risk factor for amyotrophic lateral sclerosis: A meta-analysis of epidemiological studies. Environ. Res. *117*, 112–119.

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

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013b). RNA-Guided Human Genome Engineering via Cas9. Science (80-.). *339*, 823–826.

Mann, D.M., Rollinson, S., Robinson, A., Bennion Callister, J., Thompson, J.C., Snowden, J.S., Gendron, T., Petrucelli, L., Masuda-Suzukake, M., Hasegawa, M., et al. (2013). Dipeptide repeat proteins are present in the p62 positive inclusions in patients with frontotemporal lobar degeneration and motor neurone disease associated with expansions in C9ORF72. Acta Neuropathol. Commun. *1*, 68.

Marat, A.L., Dokainish, H., and McPherson, P.S. (2011). DENN domain proteins: Regulators of Rab GTPases. J. Biol. Chem. *286*, 13791–13800.

Martier, R., Liefhebber, J.M., Miniarikova, J., van der Zon, T., Snapper, J., Kolder, I., Petry, H., van Deventer, S.J., Evers, M.M., and Konstantinova, P. (2019a). Artificial MicroRNAs Targeting C9orf72 Can Reduce Accumulation of Intra-nuclear Transcripts in ALS and FTD Patients. Mol. Ther. - Nucleic Acids *14*, 593–608.

Martier, R., Liefhebber, J.M., García-Osta, A., Miniarikova, J., Cuadrado-Tejedor, M., Espelosin, M., Ursua, S., Petry, H., van Deventer, S.J., Evers, M.M., et al. (2019b). Targeting RNA-Mediated Toxicity in C9orf72 ALS and/or FTD by RNAi-Based Gene Therapy. Mol. Ther. - Nucleic Acids *16*, 26–37.

Maruyama, H., Morino, H., Ito, H., Izumi, Y., Kato, H., Watanabe, Y., Kinoshita, Y., Kamada, M., Nodera, H., Suzuki, H., et al. (2010). Mutations of optineurin in amyotrophic lateral sclerosis. Nature *465*, 223–226.

Matsuzaki, Y., Konno, A., Mochizuki, R., Shinohara, Y., Nitta, K., Okada, Y., and Hirai, H. (2018). Intravenous administration of the adeno-associated virus-PHP.B capsid fails to upregulate transduction efficiency in the marmoset brain. Neurosci. Lett. *665*, 182–188.

Matsuzaki, Y., Tanaka, M., Hakoda, S., Masuda, T., Miyata, R., Konno, A., and Hirai, H. (2019). Neurotropic Properties of AAV-PHP.B Are Shared among Diverse Inbred Strains of Mice. Mol. Ther. *27*, 700–704.

May, S., Hornburg, D., Schludi, M.H., Arzberger, T., Rentzsch, K., Schwenk, B.M., Grässer, F.A., Mori, K., Kremmer, E., Banzhaf-Strathmann, J., et al. (2014). C9orf72 FTLD/ALS-associated Gly-Ala dipeptide repeat proteins cause neuronal toxicity and Unc119 sequestration. Acta Neuropathol. *128*, 485–503.

McCampbell, A., Cole, T., Wegener, A.J., Tomassy, G.S., Setnicka, A., Farley, B.J., Schoch, K.M., Hoye, M.L., Shabsovich, M., Sun, L., et al. (2018). Antisense oligonucleotides extend survival and reverse decrement in muscle response in ALS models. J. Clin. Invest. *128*, 3558–3567.

McCarty, D.M., Monahan, P.E., and Samulski, R.J. (2001). Self-complementary recombinant adeno-associated virus (scAAV) vectors promote efficient transduction independently of DNA synthesis. Gene Ther. *8*, 1248–1254.

McCown, T.J., Xiao, X., Li, J., Breese, G.R., and Samulski, R.J. (1996). Differential and persistent expression patterns of CNS gene transfer by an adeno-associated virus (AAV) vector. Brain Res. *713*, 99–107.

McMahon, M.A., and Cleveland, D.W. (2016). Gene therapy: Gene-editing therapy for neurological disease. Nat. Rev. Neurol. *13*, 7–9.

McMillan, C.T., Russ, J., Wood, E.M., Irwin, D.J., Grossman, M., McCluskey, L., Elman, L., Van Deerlin, V., and Lee, E.B. (2015). C9orf72 promoter hypermethylation is neuroprotective. Neurology *84*, 1622–1630.

Mehta, P., Kaye, W., Raymond, J., Wu, R., Larson, T., Punjani, R., Heller, D., Cohen, J., Peters, T., Muravov, O., et al. (2018a). Prevalence of amyotrophic lateral sclerosis - United States, 2014. Morb. Mortal. Wkly. Rep. *67*, 216–218.

Mehta, P., Kaye, W., Raymond, J., Punjani, R., Larson, T., Cohen, J., Muravov, O., and Horton, K. (2018b). Prevalence of amyotrophic lateral sclerosis — United States, 2015. Morb. Mortal. Wkly. Rep. *67*, 1285–1289.

Mendell, J.R., Al-Zaidy, S., Shell, R., Arnold, W.D., Rodino-Klapac, L.R., Prior, T.W., Lowes, L., Alfano, L., Berry, K., Church, K., et al. (2017). Single-Dose Gene-Replacement Therapy for Spinal Muscular Atrophy. N. Engl. J. Med. *377*, 1713–1722.

Meyer, K., Ferraiuolo, L., Schmelzer, L., Braun, L., McGovern, V., Likhite, S., Michels, O.,

Govoni, A., Fitzgerald, J., Morales, P., et al. (2015). Improving Single Injection CSF Delivery of AAV9-mediated Gene Therapy for SMA: A Dose–response Study in Mice and Nonhuman Primates. Mol. Ther. *23*, 477–487.

Miller, T., Cudkowicz, M., Shaw, P.J., Andersen, P.M., Atassi, N., Bucelli, R.C., Genge, A., Glass, J., Ladha, S., Ludolph, A.L., et al. (2020). Phase 1–2 Trial of Antisense Oligonucleotide Tofersen for *SOD1* ALS. N. Engl. J. Med. *383*, 109–119.

Miller, T.M., Pestronk, A., David, W., Rothstein, J., Simpson, E., Appel, S.H., Andres, P.L., Mahoney, K., Allred, P., Alexander, K., et al. (2013). An antisense oligonucleotide against SOD1 delivered intrathecally for patients with SOD1 familial amyotrophic lateral sclerosis: a phase 1, randomised, first-in-man study. Lancet Neurol. *12*, 435–442.

Mis, M.S.C., Brajkovic, S., Tafuri, F., Bresolin, N., Comi, G.P., and Corti, S. (2017). Development of Therapeutics for C9ORF72 ALS/FTD-Related Disorders. Mol. Neurobiol. *54*, 4466–4476.

Mishra, A., Ferrari, R., Heutink, P., Hardy, J., Pijnenburg, Y., Posthuma, D., and International FTD-Genomics Consortium (2017). Gene-based association studies report genetic links for clinical subtypes of frontotemporal dementia. Brain *140*, 1437–1446.

Mizielinska, S., and Isaacs, A.M. (2014). C9orf72 amyotrophic lateral sclerosis and frontotemporal dementia. Curr. Opin. Neurol. *27*, 515–523.

Mizielinska, S., Gronke, S., Niccoli, T., Ridler, C.E., Clayton, E.L., Devoy, A., Moens, T., Norona, F.E., Woollacott, I.O.C., Pietrzyk, J., et al. (2014). C9orf72 repeat expansions cause neurodegeneration in Drosophila through arginine-rich proteins. Science (80-.). *345*, 1192–1194.

Mizushima, N., and Komatsu, M. (2011). Autophagy: Renovation of cells and tissues. Cell *147*, 728–741.

Mockford, C., Jenkinson, C., and Fitzpatrick, R. (2006). A review: Carers, MND and service provision. Amyotroph. Lateral Scler. *7*, 132–141.

Moens, T.G., Niccoli, T., Wilson, K.M., Atilano, M.L., Birsa, N., Gittings, L.M., Holbling, B. V., Dyson, M.C., Thoeng, A., Neeves, J., et al. (2019). C9orf72 arginine-rich dipeptide proteins interact with ribosomal proteins in vivo to induce a toxic translational arrest that is

rescued by eIF1A. Acta Neuropathol. *137*, 487–500.

Montague, T.G., Cruz, J.M., Gagnon, J.A., Church, G.M., and Valen, E. (2014). CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. Nucleic Acids Res. *42*, 401–407.

Montagutelli, X. (2000). Effect of the Genetic Background on the Phenotype of Mouse Mutations. J. Am. Soc. Nephrol. *11*, S101–S105.

Mori, K., Weng, S.-M., Arzberger, T., May, S., Rentzsch, K., Kremmer, E., Schmid, B., Kretzschmar, H.A., Cruts, M., Van Broeckhoven, C., et al. (2013a). The C9orf72 GGGGCC Repeat Is Translated into Aggregating Dipeptide-Repeat Proteins in FTLD/ALS. Science (80-.). *339*, 1335–1338.

Mori, K., Arzberger, T., Grässer, F.A., Gijselinck, I., May, S., Rentzsch, K., Weng, S.-M., Schludi, M.H., van der Zee, J., Cruts, M., et al. (2013b). Bidirectional transcripts of the expanded C9orf72 hexanucleotide repeat are translated into aggregating dipeptide repeat proteins. Acta Neuropathol. *126*, 881–893.

Mori, K., Lammich, S., Mackenzie, I.R.A., Forné, I., Zilow, S., Kretzschmar, H., Edbauer, D., Janssens, J., Kleinberger, G., Cruts, M., et al. (2013c). HnRNP A3 binds to GGGGCC repeats and is a constituent of p62-positive/TDP43-negative inclusions in the hippocampus of patients with C9orf72 mutations. Acta Neuropathol. *125*, 413–423.

Morita, M., Al-Chalabi, A., Andersen, P.M., Hosler, B., Sapp, P., Englund, E., Mitchell, J.E., Habgood, J.J., de Belleroche, J., Xi, J., et al. (2006). A locus on chromosome 9p confers susceptibility to ALS and frontotemporal dementia. Neurology *66*, 839–844.

Mourik, J.C., Rosso, S.M., Niermeijer, M.F., Duivenvoorden, H.J., Van Swieten, J.C., and Tibben, A. (2004). Frontotemporal dementia: Behavioral symptoms and caregiver distress. Dement. Geriatr. Cogn. Disord. *18*, 299–306.

Mueller, C., Berry, J.D., McKenna-Yasek, D.M., Gernoux, G., Owegi, M.A., Pothier, L.M., Douthwright, C.L., Gelevski, D., Luppino, S.D., Blackwood, M., et al. (2020). *SOD1* Suppression with Adeno-Associated Virus and MicroRNA in Familial ALS. N. Engl. J. Med. *383*, 151–158.

Mulcahy, P.J., Binny, C., Muszynski, B., Karyka, E., and Azzouz, M. (2015). Adeno-

Associated Vectors for Gene Delivery to the Nervous System. pp. 1–22.

Münch, C., Sedlmeier, R., Meyer, T., Homberg, V., Sperfeld, A.D., Kurt, A., Prudlo, J., Peraus, G., Hanemann, C.O., Stumm, G., et al. (2004). Point mutations of the p150 subunit of dynactin (DCTN1) gene in ALS. Neurology *63*, 724–726.

Murlidharan, G., Samulski, R.J., and Asokan, A. (2014). Biology of adeno-associated viral vectors in the central nervous system. Front. Mol. Neurosci. *7*, 76.

Murphy, N.A., Arthur, K.C., Tienari, P.J., Houlden, H., Chiò, A., and Traynor, B.J. (2017). Age-related penetrance of the C9orf72 repeat expansion. Sci. Rep. *7*, 2116.

Naldini, L. (2015). Gene therapy returns to centre stage. Nature *526*, 351–360.

Nanou, A., and Azzouz, M. (2009). Gene therapy for neurodegenerative diseases based on lentiviral vectors. Prog. Brain Res. *175*, 187–200.

Nelles, D.A., Fang, M.Y., O'Connell, M.R., Xu, J.L., Markmiller, S.J., Doudna, J.A., and Yeo, G.W. (2016). Programmable RNA Tracking in Live Cells with CRISPR/Cas9. Cell *165*, 488–496.

Nelson, C.E., Hakim, C.H., Ousterout, D.G., Thakore, P.I., Moreb, E.A., Rivera, R.M.C., Madhavan, S., Pan, X., Ran, F.A., Yan, W.X., et al. (2016). In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy. Science (80-.). *351*, 403–407.

Nelson, E.J.R., Tuschong, L.M., Hunter, M.J., Bauer, T.R., Burkholder, T.H., and Hickstein, D.D. (2010). Lentiviral vectors incorporating a human elongation factor 1α promoter for the treatment of canine leukocyte adhesion deficiency. Gene Ther. *17*, 672–677.

Neumann, M., Sampathu, D.M., Kwong, L.K., Truax, A.C., Micsenyi, M.C., Chou, T.T., Bruce, J., Schuck, T., Grossman, M., Clark, C.M., et al. (2006). Ubiquitinated TDP-43 in Frontotemporal Lobar Degeneration and Amyotrophic Lateral Sclerosis. Science (80-.). *314*, 130–133.

Nguyen, L., Montrasio, F., Pattamatta, A., Tusi, S.K., Bardhi, O., Meyer, K.D., Hayes, L., Nakamura, K., Banez-Coronel, M., Coyne, A., et al. (2020). Antibody Therapy Targeting RAN Proteins Rescues C9 ALS/FTD Phenotypes in C9orf72 Mouse Model. Neuron *105*, 645-662.e11.

Nicholson, A.M., and Rademakers, R. (2016). What we know about TMEM106B in neurodegeneration. Acta Neuropathol. *132*, 639–651.

Nicolas, A., Kenna, K., Renton, A.E., Ticozzi, N., Faghri, F., Chia, R., Dominov, J.A., Kenna, B.J., Nalls, M.A., Keagle, P., et al. (2018). Genome-wide Analyses Identify KIF5A as a Novel ALS Gene. Neuron *97*, 1268-1283.e6.

Nishimura, A.L., Mitne-Neto, M., Silva, H.C.A., Oliveira, J.R.M., Vainzof, M., and Zatz, M. (2004a). A novel locus for late onset amyotrophic lateral sclerosis/motor neurone disease variant at 20q13. J. Med. Genet. *41*, 315–320.

Nishimura, A.L., Mitne-Neto, M., Silva, H.C.A., Richieri-Costa, A., Middleton, S., Cascio, D., Kok, F., Oliveira, J.R.M., Gillingwater, T., Webb, J., et al. (2004b). A mutation in the vesicle-trafficking protein VAPB causes late-onset spinal muscular atrophy and amyotrophic lateral sclerosis. Am. J. Hum. Genet. *75*, 822–831.

Nishiyama, J., Mikuni, T., and Yasuda, R. (2017). Virus-Mediated Genome Editing via Homology-Directed Repair in Mitotic and Postmitotic Cells in Mammalian Brain. Neuron *96*, 755-768.e5.

Nissim, L., Perli, S.D., Fridkin, A., Perez-Pinera, P., and Lu, T.K. (2014). Multiplexed and Programmable Regulation of Gene Networks with an Integrated RNA and CRISPR/Cas Toolkit in Human Cells. Mol. Cell *54*, 698–710.

Nordin, A., Akimoto, C., Wuolikainen, A., Alstermark, H., Jonsson, P., Birve, A., Marklund, S.L., Graffmo, K.S., Forsberg, K., Brännström, T., et al. (2015). Extensive size variability of the GGGGCC expansion in C9orf72 in both neuronal and non-neuronal tissues in 18 patients with ALS or FTD. Hum. Mol. Genet. *24*, 3133–3142.

O'Connell, M.R., Oakes, B.L., Sternberg, S.H., East-Seletsky, A., Kaplan, M., and Doudna, J.A. (2014). Programmable RNA recognition and cleavage by CRISPR/Cas9. Nature *516*, 263–266.

O'Connor, D.M., and Boulis, N.M. (2015). Gene therapy for neurodegenerative diseases. Trends Mol. Med. *21*, 504–512.

O'Rourke, J.G., Bogdanik, L., Muhammad, A.K.M.G., Gendron, T.F., Kim, K.J., Austin, A., Cady, J., Liu, E.Y., Zarrow, J., Grant, S., et al. (2015). C9orf72 BAC Transgenic Mice

Display Typical Pathologic Features of ALS/FTD. Neuron *88*, 892–901.

O'Rourke, J.G., Bogdanik, L., Yanez, A., Lall, D., Wolf, A.J., Muhammad, A.K.M.G., Ho, R., Carmona, S., Vit, J.P., Zarrow, J., et al. (2016). C9orf72 is required for proper macrophage and microglial function in mice. Science (80-.). *351*, 1324–1329.

Ohki, Y., Wenninger-Weinzierl, A., Hruscha, A., Asakawa, K., Kawakami, K., Haass, C., Edbauer, D., and Schmid, B. (2017). Glycine-alanine dipeptide repeat protein contributes to toxicity in a zebrafish model of C9orf72 associated neurodegeneration. Mol. Neurodegener. *12*, 1–11.

Oliveros, J.C., Franch, M., Tabas-Madrid, D., San-León, D., Montoliu, L., Cubas, P., and Pazos, F. (2016). Breaking-Cas—interactive design of guide RNAs for CRISPR-Cas experiments for ENSEMBL genomes. Nucleic Acids Res. *44*, W267–W271.

Olney, N.T., Spina, S., and Miller, B.L. (2017). Frontotemporal Dementia. Neurol. Clin. *35*, 339–374.

Olszewska, D.A., Lonergan, R., Fallon, E.M., and Lynch, T. (2016). Genetics of Frontotemporal Dementia. Curr. Neurol. Neurosci. Rep. *16*, 107.

Orlacchio, A., Babalini, C., Borreca, A., Patrono, C., Massa, R., Basaran, S., Munhoz, R., Rogaeva, E., St George-Hyslop, P., Bernardi, G., et al. (2010). SPATACSIN mutations cause autosomal recessive juvenile amyotrophic lateral sclerosis. Brain *133*, 591–598.

Osten, P., Dittgen, T., and Licznerski, P. (2006). Lentivirus-Based Genetic Manipulations in Neurons In Vivo (CRC Press/Taylor & Francis).

Osterlehner, A., Simmeth, S., and Göpfert, U. (2011). Promoter methylation and transgene copy numbers predict unstable protein production in recombinant chinese hamster ovary cell lines. Biotechnol. Bioeng. *108*, 2670–2681.

Ousterout, D.G., Kabadi, A.M., Thakore, P.I., Majoros, W.H., Reddy, T.E., and Gersbach, C.A. (2015). Multiplex CRISPR/Cas9-based genome editing for correction of dystrophin mutations that cause Duchenne muscular dystrophy. Nat. Commun. *6*, 6244.

Pacak, C.A., Mah, C.S., Thattaliyath, B.D., Conlon, T.J., Lewis, M.A., Cloutier, D.E., Zolotukhin, I., Tarantal, A.F., and Byrne, B.J. (2006). Recombinant Adeno-Associated Virus Serotype 9 Leads to Preferential Cardiac Transduction In Vivo. Circ. Res. *99*, e3–e9.

Paganoni, S., Macklin, E.A., Lee, A., Murphy, A., Chang, J., Zipf, A., Cudkowicz, M., and Atassi, N. (2014). Diagnostic timelines and delays in diagnosing amyotrophic lateral sclerosis (ALS). Amyotroph. Lateral Scler. Frontotemporal Degener. *15*, 453–456.

Palese, F., Sartori, A., Verriello, L., Ros, S., Passadore, P., Manganotti, P., Barbone, F., and Pisa, F.E. (2019). Epidemiology of amyotrophic lateral sclerosis in Friuli-Venezia Giulia, North-Eastern Italy, 2002–2014: a retrospective population-based study. Amyotroph. Lateral Scler. Front. Degener. *20*, 90–99.

Park, J., Childs, L., Kim, D., Hwang, G.H., Kim, S., Kim, S.T., Kim, J.S., and Bae, S. (2017). Digenome-seq web tool for profiling CRISPR specificity. Nat. Methods *14*, 548–549.

Parkinson, N., Ince, P.G., Smith, M.O., Highley, R., Skibinski, G., Andersen, P.M., Morrison, K.E., Pall, H.S., Hardiman, O., Collinge, J., et al. (2006). ALS phenotypes with mutations in CHMP2B (charged multivesicular body protein 2B). Neurology *67*, 1074–1077.

Pattanayak, V., Lin, S., Guilinger, J.P., Ma, E., Doudna, J.A., and Liu, D.R. (2013). High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. Nat. Biotechnol. *31*, 839–843.

Peters, O.M., Cabrera, G.T., Tran, H., Gendron, T.F., McKeon, J.E., Metterville, J., Weiss, A., Wightman, N., Salameh, J., Kim, J., et al. (2015). Human C9ORF72 Hexanucleotide Expansion Reproduces RNA Foci and Dipeptide Repeat Proteins but Not Neurodegeneration in BAC Transgenic Mice. Neuron *88*, 902–909.

Piccolo, P., and Brunetti-Pierri, N. (2015). Gene therapy for inherited diseases of liver metabolism. Hum. Gene Ther. *26*, 186–192.

Pick, A. (1892). Über die Beziehungen der senilen Hirnatrophie zur Aphasie. Prager Medizinische Wochenschrift *17*, 165–167.

Pinto, B.S., Saxena, T., Oliveira, R., Méndez-Gómez, H.R., Cleary, J.D., Denes, L.T., McConnell, O., Arboleda, J., Xia, G., Swanson, M.S., et al. (2017). Impeding Transcription of Expanded Microsatellite Repeats by Deactivated Cas9. Mol. Cell *68*, 479-490.e5.

Pitzer, C., Kruger, C., Plaas, C., Kirsch, F., Dittgen, T., Muller, R., Laage, R., Kastner, S., Suess, S., Spoelgen, R., et al. (2008). Granulocyte-colony stimulating factor improves outcome in a mouse model of amyotrophic lateral sclerosis. Brain *131*, 3335–3347.

Portet, F., Cadilhac, C., Touchon, J., and Camu, W. (2001). Cognitive impairment in motor neuron disease with bulbar onset. Amyotroph. Lateral Scler. Other Motor Neuron Disord. *2*, 23–29.

Pottier, C., Ravenscroft, T.A., Sanchez-Contreras, M., and Rademakers, R. (2016). Genetics of FTLD: overview and what else we can expect from genetic studies. J. Neurochem. *138*, 32–53.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell *152*, 1173–1183.

Ralph, G.S., Radcliffe, P.A., Day, D.M., Carthy, J.M., Leroux, M.A., Lee, D.C.P., Wong, L.-F., Bilsland, L.G., Greensmith, L., Kingsman, S.M., et al. (2005). Silencing mutant SOD1 using RNAi protects against neurodegeneration and extends survival in an ALS model. Nat. Med. *11*, 429–433.

Ramesh, N., and Pandey, U.B. (2017). Autophagy dysregulation in ALS: When protein aggregates get out of hand. Front. Mol. Neurosci. *10*, 263.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013a). Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. *8*, 2281–2308.

Ran, F.A., Hsu, P.D., Lin, C.-Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y., et al. (2013b). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell *154*, 1380–1389.

Ran, F.A., Cong, L., Yan, W.X., Scott, D.A., Gootenberg, J.S., Kriz, A.J., Zetsche, B., Shalem, O., Wu, X., Makarova, K.S., et al. (2015). In vivo genome editing using Staphylococcus aureus Cas9. Nature *520*, 186–191.

Raoul, C., Abbas-Terki, T., Bensadoun, J.-C., Guillot, S., Haase, G., Szulc, J., Henderson, C.E., and Aebischer, P. (2005). Lentiviral-mediated silencing of SOD1 through RNA interference retards disease onset and progression in a mouse model of ALS. Nat. Med. *11*, 423–428.

Rascovsky, K., Hodges, J.R., Knopman, D., Mendez, M.F., Kramer, J.H., Neuhaus, J., van Swieten, J.C., Seelaar, H., Dopper, E.G.P., Onyike, C.U., et al. (2011). Sensitivity of revised

diagnostic criteria for the behavioural variant of frontotemporal dementia. Brain *134*, 2456–2477.

Ratnavalli, E., Brayne, C., Dawson, K., and Hodges, J.R. (2002). The prevalence of frontotemporal dementia. Neurology *58*, 1615–1621.

Raux, G., Gantier, R., Thomas-Anterion, C., Boulliat, J., Verpillat, P., Hannequin, D., Brice, A., Frebourg, T., and Campion, D. (2000). Dementia with prominent frontotemporal features associated with L113P presenilin 1 mutation. Neurology *55*, 1577–1578.

Ravindra Kumar, S., Miles, T.F., Chen, X., Brown, D., Dobreva, T., Huang, Q., Ding, X., Luo, Y., Einarsson, P.H., Greenbaum, A., et al. (2020). Multiplexed Cre-dependent selection yields systemic AAVs for targeting distinct brain cell types. Nat. Methods *17*, 541–550.

Rembach, A., Turner, B.J., Bruce, S., Cheah, I.K., Scott, R.L., Lopes, E.C., Zagami, C.J., Beart, P.M., Cheung, N.S., Langford, S.J., et al. (2004). Antisense peptide nucleic acid targeting GluR3 delays disease onset and progression in the SOD1 G93A mouse model of familial ALS. J. Neurosci. Res. *77*, 573–582.

Renton, A.E., Majounie, E., Waite, A., Simón-Sánchez, J., Rollinson, S., Gibbs, J.R., Schymick, J.C., Laaksovirta, H., van Swieten, J.C., Myllykangas, L., et al. (2011). A Hexanucleotide Repeat Expansion in C9ORF72 Is the Cause of Chromosome 9p21-Linked ALS-FTD. Neuron *72*, 257–268.

Renton, A.E., Chiò, A., and Traynor, B.J. (2014). State of play in amyotrophic lateral sclerosis genetics. Nat. Neurosci. *17*, 17–23.

Riedijk, S.R., De Vugt, M.E., Duivenvoorden, H.J., Niermeijer, M.F., Van Swieten, J.C., Verhey, F.R.J., and Tibben, A. (2006). Caregiver burden, health-related quality of life and coping in dementia caregivers: A comparison of frontotemporal dementia and Alzheimer's disease. Dement. Geriatr. Cogn. Disord. *22*, 405–412.

Robinson, K.M., Lacey, S.C., Grugan, P., Glosser, G., Grossman, M., and McCluskey, L.F. (2006). Cognitive functioning in sporadic amyotrophic lateral sclerosis: A six month longitudinal study. J. Neurol. Neurosurg. Psychiatry *77*, 668–670.

Rohrer, J.D., Guerreiro, R., Vandrovcova, J., Uphill, J., Reiman, D., Beck, J., Isaacs, A.M., Authier, A., Ferrari, R., Fox, N.C., et al. (2009). The heritability and genetics of

frontotemporal lobar degeneration. Neurology *73*, 1451–1456.

Rohrer, J.D., Isaacs, A.M., Mizielinska, S., Mead, S., Lashley, T., Wray, S., Sidle, K., Fratta, P., Orrell, R.W., Hardy, J., et al. (2015). C9orf72 expansions in frontotemporal dementia and amyotrophic lateral sclerosis. Lancet Neurol. *14*, 291–301.

Rosen, D.R., Siddique, T., Patterson, D., Figlewicz, D.A., Sapp, P., Hentati, A., Donaldson, D., Goto, J., O'Regan, J.P., Deng, H.X., et al. (1993). Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. Nature *362*, 59–62.

Rosenbaum, M.D., VandeWoude, S., and Bielefeldt-Ohmann, H. (2007). Sudden onset of mortality within a colony of FVB/n mice. Lab Anim. (NY). *36*, 15.

Rowland, L.P. (2001). How amyotrophic lateral sclerosis got its name: the clinical-pathologic genius of Jean-Martin Charcot. Arch. Neurol. *58*, 512–515.

Russ, J., Liu, E.Y., Wu, K., Neal, D., Suh, E.R., Irwin, D.J., McMillan, C.T., Harms, M.B., Cairns, N.J., Wood, E.M., et al. (2015). Hypermethylation of repeat expanded C9orf72 is a clinical and molecular disease modifier. Acta Neuropathol. *129*, 39–52.

Ryan, D.E., Taussig, D., Steinfeld, I., Phadnis, S.M., Lunstad, B.D., Singh, M., Vuong, X., Okochi, K.D., McCaffrey, R., Olesiak, M., et al. (2018). Improving CRISPR-Cas specificity with chemical modifications in single-guide RNAs. Nucleic Acids Res. *46*, 792–803.

S. Hersh, D., S. Wadajkar, A., B. Roberts, N., G. Perez, J., P. Connolly, N., Frenkel, V., A. Winkles, J., F. Woodworth, G., and J. Kim, A. (2016). Evolving Drug Delivery Strategies to Overcome the Blood Brain Barrier. Curr. Pharm. Des. *22*, 1177–1193.

Saadatpour, Z., Bjorklund, G., Chirumbolo, S., Alimohammadi, M., Ehsani, H., Ebrahiminejad, H., Pourghadamyari, H., Baghaei, B., Mirzaei, H.R., Sahebkar, A., et al. (2016). Molecular imaging and cancer gene therapy. Cancer Gene Ther.

Salameh, J., Brown, R., and Berry, J. (2015). Amyotrophic Lateral Sclerosis: Review. Semin. Neurol. *35*, 469–476.

Sanjana, N.E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. Nat. Methods *11*, 783–784.

Sapp, P.C., Hosler, B.A., McKenna-Yasek, D., Chin, W., Gann, A., Genise, H., Gorenstein,

J., Huang, M., Sailer, W., Scheffler, M., et al. (2003). Identification of two novel loci for dominantly inherited familial amyotrophic lateral sclerosis. Am. J. Hum. Genet. *73*, 397–403.

Saraiva, J., Nobre, R.J., and Pereira de Almeida, L. (2016). Gene therapy for the CNS using AAVs: The impact of systemic delivery by AAV9. J. Control. Release *241*, 94–109.

Sareen, D., O'Rourke, J.G., Meera, P., Muhammad, A.K.M.G., Grant, S., Simpkinson, M., Bell, S., Carmona, S., Ornelas, L., Sahabian, A., et al. (2013). Targeting RNA foci in iPSC-derived motor neurons from ALS patients with a C9ORF72 repeat expansion. Sci. Transl. Med. *5*, 208ra149.

Sariyer, I.K. (2013). Transfection of neuronal cultures. Methods Mol. Biol. *1078*, 133–139.

Scarrott, J.M., Herranz-Martín, S., Alrafiah, A.R., Shaw, P.J., and Azzouz, M. (2015). Current developments in gene therapy for amyotrophic lateral sclerosis. Expert Opin. Biol. Ther. *15*, 935–947.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: An open-source platform for biological-image analysis. Nat. Methods *9*, 676–682.

Schludi, M.H., Becker, L., Garrett, L., Gendron, T.F., Zhou, Q., Schreiber, F., Popper, B., Dimou, L., Strom, T.M., Winkelmann, J., et al. (2017). Spinal poly-GA inclusions in a C9orf72 mouse model trigger motor deficits and inflammation without neuron loss. Acta Neuropathol. *134*, 241–254.

Schoch, K.M., and Miller, T.M. (2017). Antisense Oligonucleotides: Translation from Mouse Models to Human Neurodegenerative Diseases. Neuron *94*, 1056–1070.

Schwank, G., Koo, B.-K., Sasselli, V., Dekkers, J.F., Heo, I., Demircan, T., Sasaki, N., Boymans, S., Cuppen, E., van der Ent, C.K., et al. (2013). Functional Repair of CFTR by CRISPR/Cas9 in Intestinal Stem Cell Organoids of Cystic Fibrosis Patients. Cell Stem Cell *13*, 653–658.

Sellier, C., Campanari, M., Julie Corbier, C., Gaucherot, A., Kolb-Cheynel, I., Oulad-Abdelghani, M., Ruffenach, F., Page, A., Ciura, S., Kabashi, E., et al. (2016). Loss of C9 ORF 72 impairs autophagy and synergizes with polyQ Ataxin-2 to induce motor neuron dysfunction and cell death . EMBO J. *35*, 1276–1297.

Selvaraj, B.T., Livesey, M.R., Zhao, C., Gregory, J.M., James, O.T., Cleary, E.M., Chouhan, A.K., Gane, A.B., Perkins, E.M., Dando, O., et al. (2018). C9ORF72 repeat expansion causes vulnerability of motor neurons to Ca2+-permeable AMPA receptor-mediated excitotoxicity. Nat. Commun. *9*, 347.

Shao, Q., Liang, C., Chang, Q., Zhang, W., Yang, M., and Chen, J.F. (2019). C9orf72 deficiency promotes motor deficits of a C9ALS/FTD mouse model in a dose-dependent manner. Acta Neuropathol. Commun. *7*, 32.

Shi, Y., Lin, S., Staats, K.A., Li, Y., Chang, W.H., Hung, S.T., Hendricks, E., Linares, G.R., Wang, Y., Son, E.Y., et al. (2018). Haploinsufficiency leads to neurodegeneration in C9ORF72 ALS/FTD human induced motor neurons. Nat. Med. *24*, 313–325.

Shibata, N., Hirano, A., Kobayashi, M., Siddique, T., Deng, H.X., Hung, W.Y., Kato, T., and Asayama, K. (1996). Intense superoxide dismutase-1 immunoreactivity in intracytoplasmic hyaline inclusions of familial amyotrophic lateral sclerosis with posterior column involvement. J. Neuropathol. Exp. Neurol. *55*, 481–490.

Shin, J.W., Kim, K.-H., Chao, M.J., Atwal, R.S., Gillis, T., MacDonald, M.E., Gusella, J.F., and Lee, J.-M. (2016). Permanent inactivation of Huntington's disease mutation by personalized allele-specific CRISPR/Cas9. Hum. Mol. Genet. *25*, 4566–4576.

Silva-Fernandes, A., Oliveira, P., Sousa, N., and Maciel, P. (2010). Motor and Behavioural Abnormalities Associated with Persistent Spontaneous Epilepsy in the fvb/n Mouse Strain. Scand. J. Lab. Anim. Sci. *37*, 213–222.

Simone, R., Balendra, R., Moens, T.G., Preza, E., Wilson, K.M., Heslegrave, A., Woodling, N.S., Niccoli, T., Gilbert-Jaramillo, J., Abdelkarim, S., et al. (2018). G-quadruplex-binding small molecules ameliorate C9orf72 FTD / ALS pathology in vitro and in vivo . EMBO Mol. Med. *10*, 22–31.

Sivasathiaseelan, H., Marshall, C.R., Agustus, J.L., Benhamou, E., Bond, R.L., Van Leeuwen, J.E.P., Hardy, C.J.D., Rohrer, J.D., and Warren, J.D. (2019). Frontotemporal Dementia: A Clinical Review. Semin. Neurol. *39*, 251–263.

Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X., and Zhang, F. (2016). Rationally engineered Cas9 nucleases with improved specificity. Science (80-.). *351*, 84–88.

Smith, B.N., Ticozzi, N., Fallini, C., Gkazi, A.S., Topp, S., Kenna, K.P., Scotter, E.L., Kost, J., Keagle, P., Miller, J.W., et al. (2014). Exome-wide rare variant analysis identifies TUBA4A mutations associated with familial ALS. Neuron *84*, 324–331.

Smith, R.A., Miller, T.M., Yamanaka, K., Monia, B.P., Condon, T.P., Hung, G., Lobsiger, C.S., Ward, C.M., McAlonis-Downes, M., Wei, H., et al. (2006). Antisense oligonucleotide therapy for neurodegenerative disease. J. Clin. Invest. *116*, 2290–2296.

Snowden, J.S., Neary, D., and Mann, D.M.A. (2002). Frontotemporal dementia. Br. J. Psychiatry *180*, 140–143.

Spillantini, M.G., Murrell, J.R., Goedert, M., Farlow, M.R., Klug, A., and Ghetti, B. (1998). Mutation in the tau gene in familial multiple system tauopathy with presenile dementia. Proc. Natl. Acad. Sci. U. S. A. *95*, 7737–7741.

Sreedharan, J., Blair, I.P., Tripathi, V.B., Hu, X., Vance, C., Rogelj, B., Ackerley, S., Durnall, J.C., Williams, K.L., Buratti, E., et al. (2008). TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. Science (80-.). *319*, 1668–1672.

Stoica, L., Todeasa, S.H., Cabrera, G.T., Salameh, J.S., ElMallah, M.K., Mueller, C., Brown, R.H., and Sena-Esteves, M. (2016). Adeno-associated virus-delivered artificial microRNA extends survival and delays paralysis in an amyotrophic lateral sclerosis mouse model. Ann. Neurol. *79*, 687–700.

Storkebaum, E., Lambrechts, D., Dewerchin, M., Moreno-Murciano, M.-P., Appelmans, S., Oh, H., Van Damme, P., Rutten, B., Man, W.Y., De Mol, M., et al. (2005). Treatment of motoneuron degeneration by intracerebroventricular delivery of VEGF in a rat model of ALS. Nat. Neurosci. *8*, 85–92.

Stroes, E.S., Nierman, M.C., Meulenberg, J.J., Franssen, R., Twisk, J., Henny, C.P., Maas, M.M., Zwinderman, A.H., Ross, C., Aronica, E., et al. (2008). Intramuscular administration of AAV1-lipoprotein lipase S447X lowers triglycerides in lipoprotein lipase-deficient patients. Arterioscler. Thromb. Vasc. Biol. *28*, 2303–2304.

Strong, M.J. (2008). The syndromes of frontotemporal dysfunction in amyotrophic lateral sclerosis. Amyotroph. Lateral Scler. *9*, 323–338.

Strong, M.J., Grace, G.M., Orange, J.B., Leeper, H.A., Menon, R.S., and Aere, C. (1999). A

prospective study of cognitive impairment in ALS. Neurology *53*, 1665–1670.

Su, Z., Zhang, Y., Gendron, T.F., Bauer, P.O., Chew, J., Yang, W.Y., Fostvedt, E., Jansen-West, K., Belzil, V. V., Desaro, P., et al. (2014). Discovery of a Biomarker and Lead Small Molecules to Target r(GGGGCC)-Associated Defects in c9FTD/ALS. Neuron *83*, 1043–1050.

Sudria-Lopez, E., Koppers, M., de Wit, M., van der Meer, C., Westeneng, H.J., Zundel, C.A.C., Youssef, S.A., Harkema, L., de Bruin, A., Veldink, J.H., et al. (2016). Full ablation of C9orf72 in mice causes immune system-related pathology and neoplastic events but no motor neuron defects. Acta Neuropathol. *132*, 145–147.

Suh, E.R., Lee, E.B., Neal, D., Wood, E.M., Toledo, J.B., Rennert, L., Irwin, D.J., McMillan, C.T., Krock, B., Elman, L.B., et al. (2015). Semi-automated quantification of C9orf72 expansion size reveals inverse correlation between hexanucleotide repeat number and disease duration in frontotemporal degeneration. Acta Neuropathol. *130*, 363–372.

Sullivan, P.M., Zhou, X., Robins, A.M., Paushter, D.H., Kim, D., Smolka, M.B., and Hu, F. (2016). The ALS/FTLD associated protein C9orf72 associates with SMCR8 and WDR41 to regulate the autophagy-lysosome pathway. Acta Neuropathol. Commun. *4*, 51.

Suzuki, M., McHugh, J., Tork, C., Shelley, B., Hayes, A., Bellantuono, I., Aebischer, P., and Svendsen, C.N. (2008). Direct Muscle Delivery of GDNF With Human Mesenchymal Stem Cells Improves Motor Neuron Survival and Function in a Rat Model of Familial ALS. Mol. Ther. *16*, 2002–2010.

Swaminathan, A., Bouffard, M., Liao, M., Ryan, S., Bennion Callister, J., Pickering-Brown, S.M., Alan, G., Armstrong, B., and Drapeau, P. (2018). Expression of C9orf72-related dipeptides impairs motor function in a vertebrate model. Hum. Mol. Genet. *27*, 1754–1762.

Swiech, L., Heidenreich, M., Banerjee, A., Habib, N., Li, Y., Trombetta, J., Sur, M., and Zhang, F. (2014). In vivo interrogation of gene function in the mammalian brain using CRISPR-Cas9. Nat. Biotechnol. *33*, 102–106.

Tabebordbar, M., Zhu, K., Cheng, J.K.W., Chew, W.L., Widrick, J.J., Yan, W.X., Maesner, C., Wu, E.Y., Xiao, R., Ran, F.A., et al. (2016). In vivo gene editing in dystrophic mouse muscle and muscle stem cells. Science (80-.). *351*, 407–411.

Tabet, R., Schaeffer, L., Freyermuth, F., Jambeau, M., Workman, M., Lee, C.-Z., Lin, C.-C., Jiang, J., Jansen-West, K., Abou-Hamdan, H., et al. (2018). CUG initiation and frameshifting enable production of dipeptide repeat proteins from ALS/FTD C9ORF72 transcripts. Nat. Commun. *9*, 152.

Takahashi, Y., Fukuda, Y., Yoshimura, J., Toyoda, A., Kurppa, K., Moritoyo, H., Belzil, V. V., Dion, P.A., Higasa, K., Doi, K., et al. (2013). Erbb4 mutations that disrupt the neuregulin-erbb4 pathway cause amyotrophic lateral sclerosis type 19. Am. J. Hum. Genet. *93*, 900–905.

Talbot, K. (2009). Motor neuron disease: The bare essentials. Pract. Neurol. *9*, 303–309.

Talbot, K., and Wood, M.J.A. (2019). Wrangling RNA: Antisense oligonucleotides for neurological disorders. Sci. Transl. Med. *11*, eaay2069.

Talbot, P.R., Goulding, P.J., Lloyd, J.J., Snowden, J.S., Neary, D., and Testa, H.J. (1995). Inter-relation between "classic" motor neuron disease and frontotemporal dementia: neuropsychological and single photon emission computed tomography study. J. Neurol. Neurosurg. Psychiatry *58*, 541–547.

Tang, D., Sheng, J., Xu, L., Yan, C., and Qi, S. (2020a). The C9orf72-SMCR8-WDR41 complex is a GAP for small GTPases. Autophagy 1–2.

Tang, D., Sheng, J., Xu, L., Zhan, X., Liu, J., Jiang, H., Shu, X., Liu, X., Zhang, T., Jiang, L., et al. (2020b). Cryo-EM structure of C9ORF72–SMCR8–WDR41 reveals the role as a GAP for Rab8a and Rab11a. Proc. Natl. Acad. Sci. U. S. A. *117*, 9876–9883.

Taskesen, E., Mishra, A., van der Sluis, S., Ferrari, R., Veldink, J.H., van Es, M.A., Smit, A.B., Posthuma, D., Pijnenburg, Y., Hernandez, D.G., et al. (2017). Susceptible genes and disease mechanisms identified in frontotemporal dementia and frontotemporal dementia with Amyotrophic Lateral Sclerosis by DNA-methylation and GWAS. Sci. Rep. *7*, 8899.

Taylor, J.P., Brown, R.H., and Cleveland, D.W. (2016). Decoding ALS: from genes to mechanism. Nature *539*, 197–206.

Tenenbaum, L., Chtarto, A., Lehtonen, E., Velu, T., Brotchi, J., and Levivier, M. (2004). Recombinant AAV-mediated gene delivery to the central nervous system. J. Gene Med. *6*, S212–S222.

The Jackson laboratory (2020). Available at https://www.jax.org/. Summary note available

at: https://www.jax.org/strain/029099.

Therrien, M., Rouleau, G.A., Dion, P.A., Parker, J.A., Rosen, D., Siddique, T., Patterson, D., Figlewicz, D., Sapp, P., Turner, M., et al. (2013). Deletion of C9ORF72 Results in Motor Neuron Degeneration and Stress Sensitivity in C. elegans. PLoS One *8*, e83450.

Thomsen, G., Likhite, S.B., Corcoran, S., Kaspar, A., Foust, K.D., Fugere, M., Braun, L., Solano, S., Kaufmann, P., Meyer, K.C., et al. (2019). Intrathecal AAV9-SOD1-shRNA Administration for Amyotrophic Lateral Sclerosis (S5.003). Neurology *92*.

Todd, P.K., and Paulson, H.L. (2010). RNA-mediated neurodegeneration in repeat expansion disorders. Ann. Neurol. *67*, 291–300.

Tosolini, A.P., and Sleigh, J.N. (2017). Motor neuron gene therapy: Lessons from spinal muscular atrophy for amyotrophic lateral sclerosis. Front. Mol. Neurosci. *10*, 405.

Tran, H., Almeida, S., Moore, J., Gendron, T.F., Chalasani, U.D., Lu, Y., Du, X., Nickerson, J.A., Petrucelli, L., Weng, Z., et al. (2015). Differential Toxicity of Nuclear RNA Foci versus Dipeptide Repeat Proteins in a Drosophila Model of C9ORF72 FTD/ALS. Neuron *87*, 1207–1214.

Tsai, S.Q., Wyvekens, N., Khayter, C., Foden, J.A., Thapar, V., Reyon, D., Goodwin, M.J., Aryee, M.J., and Joung, J.K. (2014). Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. Nat. Biotechnol. *32*, 569–576.

Udd, B., and Krahe, R. (2012). The myotonic dystrophies: Molecular, clinical, and therapeutic challenges. Lancet Neurol. *11*, 891–905.

Ugolino, J., Ji, Y.J., Conchina, K., Chu, J., Nirujogi, R.S., Pandey, A., Brady, N.R., Hamacher-Brady, A., and Wang, J. (2016). Loss of C9orf72 Enhances Autophagic Activity via Deregulated mTOR and TFEB Signaling. PLoS Genet. *12*, e1006443.

Valera, E., and Masliah, E. (2013). Immunotherapy for neurodegenerative diseases: Focus on α- synucleinopathies. Pharmacol. Ther. *138*, 311–322.

Valori, C.F., Ning, K., Wyles, M., Mead, R.J., Grierson, A.J., Shaw, P.J., and Azzouz, M. (2010). Systemic delivery of scAAV9 expressing SMN prolongs survival in a model of spinal muscular atrophy. Sci. Transl. Med. *2*, 35ra42.

Vance, C. (2006). Familial amyotrophic lateral sclerosis with frontotemporal dementia is linked to a locus on chromosome 9p13.2-21.3. Brain *129*, 868–876.

Vance, C., Rogelj, B., Hortobágyi, T., De Vos, K.J., Nishimura, A.L., Sreedharan, J., Hu, X., Smith, B., Ruddy, D., Wright, P., et al. (2009). Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. Science (80-.). *323*, 1208–1211.

Vora, S., Cheng, J., Xiao, R., Vandusen, N.J., Quintino, L., Pu, W.T., Vandenberghe, L.H., Chavez, A., and Church, G. Rational design of a compact CRISPR--Cas9 activator for AAV--mediated delivery. Biorxiv.

Wada-Isoe, K., Ito, S., Adachi, T., Yamawaki, M., Nakashita, S., Kusumi, M., Hiroe, Y., Takada, T., Watanabe, K., Hikasa, C., et al. (2012). Epidemiological Survey of Frontotemporal Lobar Degeneration in Tottori Prefecture, Japan. Dement. Geriatr. Cogn. Dis. Extra *2*, 381–386.

Waite, A.J., Bäumer, D., East, S., Neal, J., Morris, H.R., Ansorge, O., and Blake, D.J. (2014). Reduced C9orf72 protein levels in frontal cortex of amyotrophic lateral sclerosis and frontotemporal degeneration brain with the C9ORF72 hexanucleotide repeat expansion. Neurobiol. Aging *35*, 1779.e5-1779.e13.

Walker, C., Herranz-Martin, S., Karyka, E., Liao, C., Lewis, K., Elsayed, W., Lukashchuk, V., Chiang, S.-C., Ray, S., Mulcahy, P.J., et al. (2017). C9orf72 expansion disrupts ATM-mediated chromosomal break repair. Nat. Neurosci. *20*, 1225–1235.

Wang, L.-J., Lu, Y.-Y., Muramatsu, S., Ikeguchi, K., Fujimoto, K., Okada, T., Mizukami, H., Matsushita, T., Hanazono, Y., Kume, A., et al. (2002). Neuroprotective effects of glial cell line-derived neurotrophic factor mediated by an adeno-associated virus vector in a transgenic animal model of amyotrophic lateral sclerosis. J. Neurosci. *22*, 6920–6928.

Wang, Y., Ou Mao, X., Xie, L., Banwait, S., Marti, H.H., Greenberg, D.A., and Jin, K. (2007). Vascular Endothelial Growth Factor Overexpression Delays Neurodegeneration and Prolongs Survival in Amyotrophic Lateral Sclerosis Mice. J. Neurosci. *27*, 304–307.

Wang, Z., Ma, H.I., Li, J., Sun, L., Zhang, J., and Xiao, X. (2003). Rapid and highly efficient transduction by double-stranded adeno-associated virus vectors in vitro and in vivo. Gene Ther. *10*, 2105–2111.

Wang, Z., Wang, S., Li, D., Zhang, Q., Li, L., Zhong, C., Liu, Y., and Huang, H. (2018). Optimized paired-sgRNA/Cas9 cloning and expression cassette triggers high-efficiency multiplex genome editing in kiwifruit. Plant Biotechnol. J. *16*, 1424–1433.

Ward, J.M., Mahler, J.F., Maronport, R.R., Sundberg, J.P., and Frederickson, R.M. (2000). Pathology of genetically engineered mice. Pathol. Genet. Eng. Mice.

Watts, G.D.J., Wymer, J., Kovach, M.J., Mehta, S.G., Mumm, S., Darvish, D., Pestronk, A., Whyte, M.P., and Kimonis, V.E. (2004). Inclusion body myopathy associated with Paget disease of bone and frontotemporal dementia is caused by mutant valosin-containing protein. Nat. Genet. *36*, 377–381.

Webster, C.P., Smith, E.F., Grierson, A.J., and De Vos, K.J. (2016a). C9orf72 plays a central role in Rab GTPase-dependent regulation of autophagy. Small GTPases *9*, 399–408.

Webster, C.P., Smith, E.F., Bauer, C.S., Moller, A., Hautbergue, G.M., Ferraiuolo, L., Myszczynska, M.A., Higginbottom, A., Walsh, M.J., Whitworth, A.J., et al. (2016b). The C9orf72 protein interacts with Rab1a and the ULK1 complex to regulate initiation of autophagy. EMBO J. *35*, 1656–1676.

Wen, X., Tan, W., Westergard, T., Krishnamurthy, K., Markandaiah, S.S., Shi, Y., Lin, S., Shneider, N.A., Monaghan, J., Pandey, U.B., et al. (2014). Antisense proline-arginine RAN dipeptides linked to C9ORF72-ALS/FTD form toxic nuclear aggregates that initiate invitro and invivo neuronal death. Neuron *84*, 1213–1225.

Wienert, B., Wyman, S.K., Richardson, C.D., Yeh, C.D., Akcakaya, P., Porritt, M.J., Morlock, M., Vu, J.T., Kazane, K.R., Watry, H.L., et al. (2019). Unbiased detection of CRISPR off-targets in vivo using DISCOVER-Seq. Science (80-.). *364*, 286–289.

Wu, C.H., Fallini, C., Ticozzi, N., Keagle, P.J., Sapp, P.C., Piotrowska, K., Lowe, P., Koppers, M., McKenna-Yasek, D., Baron, D.M., et al. (2012). Mutations in the profilin 1 gene cause familial amyotrophic lateral sclerosis. Nature *488*, 499–503.

Wu, Y., Liang, D., Wang, Y., Bai, M., Tang, W., Bao, S., Yan, Z., Li, D., and Li, J. (2013). Correction of a Genetic Disease in Mouse via Use of CRISPR-Cas9. Cell Stem Cell *13*, 659–662.

Wu, Z., Yang, H., and Colosi, P. (2010). Effect of genome size on AAV vector packaging.

Mol. Ther. *18*, 80–86.

Xi, Z., Zinman, L., Moreno, D., Schymick, J., Liang, Y., Sato, C., Zheng, Y., Ghani, M., Dib, S., Keith, J., et al. (2013). Hypermethylation of the CpG island near the G4C2 repeat in ALS with a C9orf72 expansion. Am. J. Hum. Genet. *92*, 981–989.

Xi, Z., Zhang, M., Bruni, A.C., Maletta, R.G., Colao, R., Fratta, P., Polke, J.M., Sweeney, M.G., Mudanohwo, E., Nacmias, B., et al. (2015). The C9orf72 repeat expansion itself is methylated in ALS and FTLD patients. Acta Neuropathol. *129*, 715–727.

Xiao-Jie, L., Hui-Ying, X., Zun-Ping, K., Jin-Lian, C., and Li-Juan, J. (2015). CRISPR-Cas9: a new and promising player in gene therapy. J. Med. Genet. *52*, 289–296.

Xiao, S., MacNair, L., McGoldrick, P., McKeever, P.M., McLean, J.R., Zhang, M., Keith, J., Zinman, L., Rogaeva, E., and Robertson, J. (2015). Isoform-specific antibodies reveal distinct subcellular localizations of C9orf72 in amyotrophic lateral sclerosis. Ann. Neurol. *78*, 568–583.

Xiao, S., MacNair, L., McLean, J., McGoldrick, P., McKeever, P., Soleimani, S., Keith, J., Zinman, L., Rogaeva, E., and Robertson, J. (2016). C9orf72 isoforms in Amyotrophic Lateral Sclerosis and Frontotemporal Lobar Degeneration. Brain Res. *1647*, 43–49.

Xie, K., Minkenberg, B., and Yang, Y. (2015). Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. Proc. Natl. Acad. Sci. U. S. A. *112*, 3570–3575.

Xu, W., and Xu, J. (2018). C9orf72 dipeptide repeats cause selective neurodegeneration and cell-autonomous excitotoxicity in Drosophila glutamatergic neurons. J. Neurosci. *38*, 7741–7752.

Xu, L., Zhao, L., Gao, Y., Xu, J., and Han, R. (2017). Empower multiplex cell and tissue-specific CRISPR-mediated gene manipulation with self-cleaving ribozymes and tRNA. Nucleic Acids Res. *45*, e28.

Xu, Z., Poidevin, M., Li, X., Li, Y., Shu, L., Nelson, D.L., Li, H., Hales, C.M., Gearing, M., Wingo, T.S., et al. (2013). Expanded GGGGCC repeat RNA associated with amyotrophic lateral sclerosis and frontotemporal dementia causes neurodegeneration. Proc. Natl. Acad. Sci. U. S. A. *110*, 7778–7783.

Xue, H.-Y., Zhang, X., Wang, Y., Xiaojie, L., Dai, W.-J., and Xu, Y. (2016). In vivo gene therapy potentials of CRISPR-Cas9. Gene Ther. *23*, 557–559.

Yamashita, S., Mita, S., Kato, S., Okado, H., Ohama, E., and Uchino, M. (2002). Effect on motor neuron survival in mutant SOD1 (G93A) transgenic mice by Bcl-2 expression using retrograde axonal transport of adenoviral vectors. Neurosci. Lett. *328*, 289–293.

Yan, J., Deng, H.X., Siddique, N., Fecto, F., Chen, W., Yang, Y., Liu, E., Donkervoort, S., Zheng, J.G., Shi, Y., et al. (2010). Frameshift and novel mutations in FUS in familial amyotrophic lateral sclerosis and ALS/dementia. Neurology *75*, 807–814.

Yang, M., Chen, L., Swaminathan, K., Herrlinger, S., Lai, F., Shiekhattar, R., and Chen, J.F. (2016). A C9ORF72/SMCR8-containing complex regulates ULK1 and plays a dual role in autophagy. Sci. Adv. *2*, e1601167.

Yang, Y., Hentati, A., Deng, H.X., Dabbagh, O., Sasaki, T., Hirano, M., Hung, W.Y., Ouahchi, K., Yan, J., Azim, A.C., et al. (2001). The gene encoding alsin, a protein with three guanine-nucleotide exchange factor domains, is mutated in a form of recessive amyotrophic lateral sclerosis. Nat. Genet. *29*, 160–165.

Yates, A., Akanni, W., Amode, M.R., Barrell, D., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Fitzgerald, S., Gil, L., et al. (2016). Ensembl 2016. Nucleic Acids Res. *44*, D710–D716.

Yates, A.D., Achuthan, P., Akanni, W., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M.R., Armean, I.M., Azov, A.G., Bennett, R., et al. (2020). Ensembl 2020. Nucleic Acids Res. *48*, D682–D688.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T.L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics *13*, 134.

Ye, L., Wang, J., Beyer, A.I., Teque, F., Cradick, T.J., Qi, Z., Chang, J.C., Bao, G., Muench, M.O., Yu, J., et al. (2014). Seamless modification of wild-type induced pluripotent stem cells to the natural CCR5Δ32 mutation confers resistance to HIV infection. Proc. Natl. Acad. Sci. *111*, 9591–9596.

Yin, H., Kanasty, R.L., Eltoukhy, A.A., Vegas, A.J., Dorkin, J.R., and Anderson, D.G.

(2014a). Non-viral vectors for gene-based therapy. Nat. Rev. Genet. *15*, 541–555.

Yin, H., Xue, W., Chen, S., Bogorad, R.L., Benedetti, E., Grompe, M., Koteliansky, V., Sharp, P.A., Jacks, T., and Anderson, D.G. (2014b). Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. Nat. Biotechnol. *32*, 551–553.

Yoshioka, S., Fujii, W., Ogawa, T., Sugiura, K., and Naito, K. (2015). Development of a mono-promoter-driven CRISPR/Cas9 system in mammalian cells. Sci. Rep. *5*, 18341.

Zamiri, B., Reddy, K., Macgregor, R.B., and Pearson, C.E. (2014). TMPyP4 Porphyrin Distorts RNA G-quadruplex Structures of the Disease-associated r(GGGGCC) *n* Repeat of the *C9orf72* Gene and Blocks Interaction of RNA-binding Proteins. J. Biol. Chem. *289*, 4653–4659.

Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M.R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Girón, C.G., et al. (2018). Ensembl 2018. Nucleic Acids Res. *46*, D754–D761.

Zhang, W.W., and Matlashewski, G. (2015). CRISPR-Cas9-mediated genome editing in Leishmania donovani. MBio *6*, e00861.

Zhang, D., Iyer, L.M., He, F., and Aravind, L. (2012). Discovery of Novel DENN Proteins: Implications for the Evolution of Eukaryotic Intracellular Membrane Structures and Human Disease. Front. Genet. *3*, 283.

Zhang, H., Yang, B., Mu, X., Ahmed, S.S., Su, Q., He, R., Wang, H., Mueller, C., Sena-Esteves, M., Brown, R., et al. (2011). Several rAAV Vectors Efficiently Cross the Blood–brain Barrier and Transduce Neurons and Astrocytes in the Neonatal Mouse Central Nervous System. Mol. Ther. *19*, 1440–1448.

Zhang, K., Donnelly, C.J., Haeusler, A.R., Grima, J.C., Machamer, J.B., Steinwald, P., Daley, E.L., Miller, S.J., Cunningham, K.M., Vidensky, S., et al. (2015). The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. Nature *525*, 56–61.

Zhang, Y.J., Jansen-West, K., Xu, Y.F., Gendron, T.F., Bieniek, K.F., Lin, W.L., Sasaguri, H., Caulfield, T., Hubbard, J., Daughrity, L., et al. (2014). Aggregation-prone c9FTD/ALS poly(GA) RAN-translated proteins cause neurotoxicity by inducing ER stress. Acta Neuropathol. *128*, 505–524.

Zhang, Y.J., Gendron, T.F., Grima, J.C., Sasaguri, H., Jansen-West, K., Xu, Y.F., Katzman, R.B., Gass, J., Murray, M.E., Shinohara, M., et al. (2016). C9ORF72 poly(GA) aggregates sequester and impair HR23 and nucleocytoplasmic transport proteins. Nat. Neurosci. *19*, 668–677.

Zhang, Y.J., Gendron, T.F., Ebbert, M.T.W., O'Raw, A.D., Yue, M., Jansen-West, K., Zhang, X., Prudencio, M., Chew, J., Cook, C.N., et al. (2018). Poly(GR) impairs protein translation and stress granule dynamics in C9orf72-associated frontotemporal dementia and amyotrophic lateral sclerosis. Nat. Med. *24*, 1136–1142.

Zhou, Q., Lehmer, C., Michaelsen, M., Mori, K., Alterauge, D., Baumjohann, D., Schludi, M.H., Greiling, J., Farny, D., Flatley, A., et al. (2017). Antibodies inhibit transmission and aggregation of C9orf72 poly- GA dipeptide repeat proteins . EMBO Mol. Med. *9*, 687–702.

Zhou, Q., Mareljic, N., Michaelsen, M., Parhizkar, S., Heindl, S., Nuscher, B., Farny, D., Czuppa, M., Schludi, C., Graf, A., et al. (2020). Active poly-GA vaccination prevents microglia activation and motor deficits in a C9orf72 mouse model. EMBO Mol. Med. *12*, e10919.

Zhu, Q., Jiang, J., Gendron, T.F., McAlonis-Downes, M., Jiang, L., Taylor, A., Diaz Garcia, S., Ghosh Dastidar, S., Rodriguez, M.J., King, P., et al. (2020). Reduced C9ORF72 function exacerbates gain of toxicity from ALS/FTD-causing repeat expansion in C9orf72. Nat. Neurosci. *23*, 615–624.

Zincarelli, C., Soltys, S., Rengo, G., and Rabinowitz, J.E. (2008). Analysis of AAV serotypes 1-9 mediated gene expression and tropism in mice after systemic injection. Mol. Ther. *16*, 1073–1080.

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

Zu, T., Gibbens, B., Doty, N.S., Gomes-Pereira, M., Huguet, A., Stone, M.D., Margolis, J., Peterson, M., Markowski, T.W., Ingram, M.A.C., et al. (2011). Non-ATG-initiated translation directed by microsatellite expansions. Proc. Natl. Acad. Sci. U. S. A. *108*, 260–265.

Zu, T., Liu, Y., Bañez-Coronel, M., Reid, T., Pletnikova, O., Lewis, J., Miller, T.M., Harms,

M.B., Falchook, A.E., Subramony, S.H., et al. (2013). RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia. Proc. Natl. Acad. Sci. U. S. A. *110*, E4968.

Zu, T., Guo, S., Bardhi, O., Ryskamp, D.A., Li, J., Khoramian Tusi, S., Engelbrecht, A., Klippel, K., Chakrabarty, P., Nguyen, L., et al. (2020). Metformin inhibits RAN translation through PKR pathway and mitigates disease in *C9orf72* ALS/FTD mice. Proc. Natl. Acad. Sci. 202005748.

(2007). Diagnosis: “Space cadet” syndrome of female FVB/n mice. Lab Anim. (NY). *36*, 16.

Appendices

Appendix 1. Plasmid sequences.

(a) Sequence of pX601::dual sgRNA cassette

cctgcaggcagctgcgcgctcgctcgctcactgaggccgcccgggcgtcgggcgacctttggtcgcccggcctcagtgagcgagc
gagcgcgcagagagggagtggccaactccatcactaggggttcctgcggcctctagactcgaggcgttgacattgattattgactagt
tattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctggctg
accgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatggg
tggagtatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatg
gcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggtgat
gcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagttt
gttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgg
gaggtctatataagcagagctctctggctaactaccggtgccaccATGGCCCCAAAGAAGAAGCGGAAGGT
CGGTATCCACGGAGTCCCAGCAGCCAAGCGGAACTACATCCTGGGCCTGGACAT
CGGCATCACCAGCGTGGGCTACGGCATCATCGACTACGAGACACGGGACGTGAT
CGATGCCGGCGTGCGGCTGTTCAAAGAGGCCAACGTGGAAAACAACGAGGGCA
GGCGGAGCAAGAGAGGCGCCAGAAGGCTGAAGCGGCGGAGGCGGCATAGAATC
CAGAGAGTGAAGAAGCTGCTGTTCGACTACAACCTGCTGACCGACCACAGCGAG
CTGAGCGGCATCAACCCCTACGAGGCCAGAGTGAAGGGCCTGAGCCAGAAGCTG
AGCGAGGAAGAGTTCTCTGCCGCCCTGCTGCACCTGGCCAAGAGAAGAGGCGTG
CACAACGTGAACGAGGTGGAAGAGGACACCGGCAACGAGCTGTCCACCAAAGA
GCAGATCAGCCGGAACAGCAAGGCCCTGGAAGAGAAATACGTGGCCGAACTGC
AGCTGGAACGGCTGAAGAAAGACGGCGAAGTGCGGGGCAGCATCAACAGATTC
AAGACCAGCGACTACGTGAAAGAAGCCAAACAGCTGCTGAAGGTGCAGAAGGC
CTACCACCAGCTGGACCAGAGCTTCATCGACACCTACATCGACCTGCTGGAAACC
CGGCGGACCTACTATGAGGGACCTGGCGAGGGCAGCCCCTTCGGCTGGAAGGAC
ATCAAAGAATGGTACGAGATGCTGATGGGCCACTGCACCTACTTCCCCGAGGAA
CTGCGGAGCGTGAAGTACGCCTACAACGCCGACCTGTACAACGCCCTGAACGAC

CTGAACAATCTCGTGATCACCAGGGACGAGAACGAGAAGCTGGAATATTACGAG
AAGTTCCAGATCATCGAGAACGTGTTCAAGCAGAAGAAGAAGCCCACCCTGAAG
CAGATCGCCAAAGAAATCCTCGTGAACGAAGAGGATATTAAGGGCTACAGAGTG
ACCAGCACCGGCAAGCCCGAGTTCACCAACCTGAAGGTGTACCACGACATCAAG
GACATTACCGCCCGGAAAGAGATTATTGAGAACGCCGAGCTGCTGGATCAGATT
GCCAAGATCCTGACCATCTACCAGAGCAGCGAGGACATCCAGGAAGAACTGACC
AATCTGAACTCCGAGCTGACCCAGGAAGAGATCGAGCAGATCTCTAATCTGAAG
GGCTATACCGGCACCCACAACCTGAGCCTGAAGGCCATCAACCTGATCCTGGAC
GAGCTGTGGCACACCAACGACAACCAGATCGCTATCTTCAACCGGCTGAAGCTG
GTGCCCAAGAAGGTGGACCTGTCCCAGCAGAAAGAGATCCCCACCACCCTGGTG
GACGACTTCATCCTGAGCCCCGTCGTGAAGAGAAGCTTCATCCAGAGCATCAAA
GTGATCAACGCCATCATCAAGAAGTACGGCCTGCCCAACGACATCATTATCGAG
CTGGCCCGCGAGAAGAACTCCAAGGACGCCCAGAAAATGATCAACGAGATGCA
GAAGCGGAACCGGCAGACCAACGAGCGGATCGAGGAAATCATCCGGACCACCG
GCAAAGAGAACGCCAAGTACCTGATCGAGAAGATCAAGCTGCACGACATGCAG
GAAGGCAAGTGCCTGTACAGCCTGGAAGCCATCCCTCTGGAAGATCTGCTGAAC
AACCCCTTCAACTATGAGGTGGACCACATCATCCCCAGAAGCGTGTCCTTCGACA
ACAGCTTCAACAACAAGGTGCTCGTGAAGCAGGAAGAAAACAGCAAGAAGGGC
AACCGGACCCCATTCCAGTACCTGAGCAGCAGCGACAGCAAGATCAGCTACGAA
ACCTTCAAGAAGCACATCCTGAATCTGGCCAAGGGCAAGGGCAGAATCAGCAAG
ACCAAGAAAGAGTATCTGCTGGAAGAACGGGACATCAACAGGTTCTCCGTGCAG
AAAGACTTCATCAACCGGAACCTGGTGGATACCAGATACGCCACCAGAGGCCTG
ATGAACCTGCTGCGGAGCTACTTCAGAGTGAACAACCTGGACGTGAAAGTGAAG
TCCATCAATGGCGGCTTCACCAGCTTTCTGCGGCGGAAGTGGAAGTTTAAGAAA
GAGCGGAACAAGGGGTACAAGCACCACGCCGAGGACGCCCTGATCATTGCCAAC
GCCGATTTCATCTTCAAAGAGTGGAAGAAACTGGACAAGGCCAAAAAAGTGATG
GAAAACCAGATGTTCGAGGAAAAGCAGGCCGAGAGCATGCCCGAGATCGAAAC
CGAGCAGGAGTACAAAGAGATCTTCATCACCCCCCACCAGATCAAGCACATTAA
GGACTTCAAGGACTACAAGTACAGCCACCGGGTGGACAAGAAGCCTAATAGAGA
GCTGATTAACGACACCCTGTACTCCACCCGGAAGGACGACAAGGGCAACACCCT
GATCGTGAACAATCTGAACGGCCTGTACGACAAGGACAATGACAAGCTGAAAAA
GCTGATCAACAAGAGCCCCGAAAAGCTGCTGATGTACCACCACGACCCCCAGAC
CTACCAGAAACTGAAGCTGATTATGGAACAGTACGGCGACGAGAAGAATCCCCT
GTACAAGTACTACGAGGAAACCGGGAACTACCTGACCAAGTACTCCAAAAAGGA

CAACGGCCCCGTGATCAAGAAGATTAAGTATTACGGCAACAAACTGAACGCCCA
TCTGGACATCACCGACGACTACCCCAACAGCAGAAACAAGGTCGTGAAGCTGTC
CCTGAAGCCCTACAGATTCGACGTGTACCTGGACAATGGCGTGTACAAGTTCGTG
ACCGTGAAGAATCTGGATGTGATCAAAAAAGAAAACTACTACGAAGTGAATAGC
AAGTGCTATGAGGAAGCTAAGAAGCTGAAGAAGATCAGCAACCAGGCCGAGTTT
ATCGCCTCCTTCTACAACAACGATCTGATCAAGATCAACGGCGAGCTGTATAGAG
TGATCGGCGTGAACAACGACCTGCTGAACCGGATCGAAGTGAACATGATCGACA
TCACCTACCGCGAGTACCTGGAAAACATGAACGACAAGAGGCCCCCCAGGATCA
TTAAGACAATCGCCTCCAAGACCCAGAGCATTAAGAAGTACAGCACAGACATTC
TGGGCAACCTGTATGAAGTGAAATCTAAGAAGCACCCTCAGATCATCAAAAAGG
GCAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGggat
cctacccatacgatgttccagattacgcttacccatacgatgttccagattacgcttacccatacgatgttccagattacgcttaagaattcc
tagagctcgctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctcccccgtgccttccttgaccctggaaggt
gccactcccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctgggggggtggggtggggc
aggacagcaaggggaggattgggaagagaatagcaggcatgctggggaggtacCTGAGGGCCTATTTCCCAT
GATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGGAATT
AATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAA
TAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATAT
GCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAG
GACGAAACACCGGAGACCACGGCAGGTCTCAGTTTTAGTACTCTGGAAACAGAA
TCTACTAAAACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGATT
TTTGCCTGAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGG
CTGTTAGAGAGATAATTGGAATTAATTTGACTGTAAACACAAAGATATTAGTACA
AAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTA
TGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCGATTTCTT
GGCTTTATATATCTTGTGGAAAGGACGAAACACCGGGTCTTCGGATCCATGCGAA
GACCTGTTTTAGTACTCTGGAAACAGAATCTACTAAAACAAGGCAAAATGCCGT
GTTTATCTCGTCAACTTGTTGGCGAGATTTTTGCggccgcaggaaccctagtgatggagttggccac
tccctctctgcgcgctcgctcgctcactgaggccgggcgaccaaaggtcgcccgacgcccgggctttgcccgggcggcctcagtga
gcgagcgagcgcgcagctgcctgcaggggcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgcatacgtcaa
agcaaccatagtacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagc
gccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctcccttta
gggttccgatttagtgctttacggcacctcgaccccaaaaaacttgatttgggtgatggttcacgtagtgggccatcgccctgatagacg
gtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcgggctattctt

tgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacg
tttacaattttatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagccccgacacccgccaacacccgctgacgcg
ccctgacgggcttgtctgctcccggcatccgcttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtc
atcaccgaaacgcgcgagacgaaagggcctcgtgatacgcctatttttataggttaatgtcatgataataatggtttcttagacgtcaggt
ggcacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgat
aaatgcttcaataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattccctttttgcggcattttgccttcctgttt
ttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacag
cggtaagatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtat
tgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatctta
cggatggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcgg
aggaccgaaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagcc
ataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctag
cttcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttatt
gctgataaatctggagccggtgagcgtggaagccgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagtta
tctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaact
gtcagaccaagtttactcatatatactttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaagatccttttgataatctcatg
accaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatccttttttctgcg
cgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggt
aactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgc
ctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagtt
accggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgag
atacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcgg
aacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgt
cgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcctggccttttgctggcct
tttgctcacatgt

(b) Sequence of pX601::EFS1a::dual sgRNA cassette

cctgcaggcagctgcgcgctcgctcgctcactgaggccgcccgggcgtcgggcgacctttggtcgcccggcctcagtgagcgagc
gagcgcgcagagagggagtggccaactccatcactaggggttcctgcggcctCTAGAGGCTCCGGTGCCCGTC

AGTGGGCAGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCG
GCAATTGATCCGGTGCCTAGAGAAGGTGGCGCGGGGTAAACTGGGAAAGTGATG
TCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGC
AGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGGCc
ATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCAA
GCGGAACTACATCCTGGGCCTGGACATCGGCATCACCAGCGTGGGCTACGGCAT
CATCGACTACGAGACACGGGACGTGATCGATGCCGGCGTGCGGCTGTTCAAAGA
GGCCAACGTGGAAAACAACGAGGGCAGGCGGAGCAAGAGAGGCGCCAGAAGGC
TGAAGCGGCGGAGGCGGCATAGAATCCAGAGAGTGAAGAAGCTGCTGTTCGACT
ACAACCTGCTGACCGACCACAGCGAGCTGAGCGGCATCAACCCCTACGAGGCCA
GAGTGAAGGGCCTGAGCCAGAAGCTGAGCGAGGAAGAGTTCTCTGCCGCCCTGC
TGCACCTGGCCAAGAGAAGAGGCGTGCACAACGTGAACGAGGTGGAAGAGGAC
ACCGGCAACGAGCTGTCCACCAAAGAGCAGATCAGCCGGAACAGCAAGGCCCTG
GAAGAGAAATACGTGGCCGAACTGCAGCTGGAACGGCTGAAGAAAGACGGCGA
AGTGCGGGGCAGCATCAACAGATTCAAGACCAGCGACTACGTGAAAGAAGCCA
AACAGCTGCTGAAGGTGCAGAAGGCCTACCACCAGCTGGACCAGAGCTTCATCG
ACACCTACATCGACCTGCTGGAAACCCGGCGGACCTACTATGAGGGACCTGGCG
AGGGCAGCCCCTTCGGCTGGAAGGACATCAAAGAATGGTACGAGATGCTGATGG
GCCACTGCACCTACTTCCCCGAGGAACTGCGGAGCGTGAAGTACGCCTACAACG
CCGACCTGTACAACGCCCTGAACGACCTGAACAATCTCGTGATCACCAGGGACG
AGAACGAGAAGCTGGAATATTACGAGAAGTTCCAGATCATCGAGAACGTGTTCA
AGCAGAAGAAGAAGCCCACCCTGAAGCAGATCGCCAAAGAAATCCTCGTGAAC
GAAGAGGATATTAAGGGCTACAGAGTGACCAGCACCGGCAAGCCCGAGTTCACC
AACCTGAAGGTGTACCACGACATCAAGGACATTACCGCCCGGAAAGAGATTATT
GAGAACGCCGAGCTGCTGGATCAGATTGCCAAGATCCTGACCATCTACCAGAGC
AGCGAGGACATCCAGGAAGAACTGACCAATCTGAACTCCGAGCTGACCCAGGAA
GAGATCGAGCAGATCTCTAATCTGAAGGGCTATACCGGCACCCACAACCTGAGC
CTGAAGGCCATCAACCTGATCCTGGACGAGCTGTGGCACACCAACGACAACCAG
ATCGCTATCTTCAACCGGCTGAAGCTGGTGCCCAAGAAGGTGGACCTGTCCCAGC
AGAAAGAGATCCCCACCACCCTGGTGGACGACTTCATCCTGAGCCCCGTCGTGA
AGAGAAGCTTCATCCAGAGCATCAAAGTGATCAACGCCATCATCAAGAAGTACG
GCCTGCCCAACGACATCATTATCGAGCTGGCCCGCGAGAAGAACTCCAAGGACG
CCCAGAAAATGATCAACGAGATGCAGAAGCGGAACCGGCAGACCAACGAGCGG
ATCGAGGAAATCATCCGGACCACCGGCAAAGAGAACGCCAAGTACCTGATCGAG

AAGATCAAGCTGCACGACATGCAGGAAGGCAAGTGCCTGTACAGCCTGGAAGCC
ATCCCTCTGGAAGATCTGCTGAACAACCCCTTCAACTATGAGGTGGACCACATCA
TCCCCAGAAGCGTGTCCTTCGACAACAGCTTCAACAACAAGGTGCTCGTGAAGC
AGGAAGAAAACAGCAAGAAGGGCAACCGGACCCCATTCCAGTACCTGAGCAGC
AGCGACAGCAAGATCAGCTACGAAACCTTCAAGAAGCACATCCTGAATCTGGCC
AAGGGCAAGGGCAGAATCAGCAAGACCAAGAAAGAGTATCTGCTGGAAGAACG
GGACATCAACAGGTTCTCCGTGCAGAAAGACTTCATCAACCGGAACCTGGTGGA
TACCAGATACGCCACCAGAGGCCTGATGAACCTGCTGCGGAGCTACTTCAGAGT
GAACAACCTGGACGTGAAAGTGAAGTCCATCAATGGCGGCTTCACCAGCTTTCT
GCGGCGGAAGTGGAAGTTTAAGAAAGAGCGGAACAAGGGGTACAAGCACCACG
CCGAGGACGCCCTGATCATTGCCAACGCCGATTTCATCTTCAAAGAGTGGAAGA
AACTGGACAAGGCCAAAAAAGTGATGGAAAACCAGATGTTCGAGGAAAAGCAG
GCCGAGAGCATGCCCGAGATCGAAACCGAGCAGGAGTACAAAGAGATCTTCATC
ACCCCCCACCAGATCAAGCACATTAAGGACTTCAAGGACTACAAGTACAGCCAC
CGGGTGGACAAGAAGCCTAATAGAGAGCTGATTAACGACACCCTGTACTCCACC
CGGAAGGACGACAAGGGCAACACCCTGATCGTGAACAATCTGAACGGCCTGTAC
GACAAGGACAATGACAAGCTGAAAAAGCTGATCAACAAGAGCCCCGAAAAGCT
GCTGATGTACCACCACGACCCCCAGACCTACCAGAAACTGAAGCTGATTATGGA
ACAGTACGGCGACGAGAAGAATCCCCTGTACAAGTACTACGAGGAAACCGGGA
ACTACCTGACCAAGTACTCCAAAAAGGACAACGGCCCCGTGATCAAGAAGATTA
AGTATTACGGCAACAAACTGAACGCCCATCTGGACATCACCGACGACTACCCCA
ACAGCAGAAACAAGGTCGTGAAGCTGTCCCTGAAGCCCTACAGATTCGACGTGT
ACCTGGACAATGGCGTGTACAAGTTCGTGACCGTGAAGAATCTGGATGTGATCA
AAAAAGAAAACTACTACGAAGTGAATAGCAAGTGCTATGAGGAAGCTAAGAAG
CTGAAGAAGATCAGCAACCAGGCCGAGTTTATCGCCTCCTTCTACAACAACGATC
TGATCAAGATCAACGGCGAGCTGTATAGAGTGATCGGCGTGAACAACGACCTGC
TGAACCGGATCGAAGTGAACATGATCGACATCACCTACCGCGAGTACCTGGAAA
ACATGAACGACAAGAGGCCCCCCAGGATCATTAAGACAATCGCCTCCAAGACCC
AGAGCATTAAGAAGTACAGCACAGACATTCTGGGCAACCTGTATGAAGTGAAAT
CTAAGAAGCACCCTCAGATCATCAAAAAGGGCAAAAGGCCGGCGGCCACGAAA
AAGGCCGGCCAGGCAAAAAAGAAAAAGggatcctacccatacgatgttccagattacgcttacccatacga
tgttccagattacgcttacccatacgatgttccagattacgcttaagaattcctagagctcgctgatcagcctcgactgtgccttctagttgc
cagccatctgttgtttgcccctcccccgtgccttccttgaccctggaaggtgccactcccactgtcctttcctaataaaatgaggaaattgc
atcgcattgtctgagtaggtgtcattctattctgggggggtggggtggggcaggacagcaagggggaggattgggaagagaatagca

ggcatgctggggaggtacCTGAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGA
TACAAGGCTGTTAGAGAGATAATTGGAATTAATTTGACTGTAAACACAAAGATA
TTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTT
TAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTT
CGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGGAGACCACGGC
AGGTCTCAGTTTTAGTACTCTGGAAACAGAATCTACTAAAACAAGGCAAAATGC
CGTGTTTATCTCGTCAACTTGTTGGCGAGATTTTTGCCTGAGGGCCTATTTCCCAT
GATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGGAATT
AATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAA
TAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATAT
GCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAG
GACGAAACACCGGGTCTTCGGATCCATGCGAAGACCTGTTTTAGTACTCTGGAAA
CAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCG
AGATTTTTGCggccgcaggaacccctagtgatggagttggccactccctctctgcgcgctcgctcgctcactgaggccggg
cgaccaaaggtcgcccgacgcccgggctttgcccgggcggcctcagtgagcgagcgagcgcgcagctgcctgcaggggcgcct
gatgcggtattttctccttacgcatctgtgcggtatttcacaccgcatacgtcaaagcaaccatagtacgcgccctgtagcggcgcattaa
gcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttc
tcgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccgatttagtgctttacggcacctcgacccca
aaaaacttgatttgggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaata
gtggactcttgttccaaactggaacaacactcaaccctatctcgggctattcttttgatttataagggattttgccgatttcggcctattggtt
aaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaattttatggtgcactctcagtacaatctgctct
gatgccgcatagttaagccagccccgacacccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggcatccgcttac
agacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgagacgaaagggcctcgtg
atacgcctatttttataggttaatgtcatgataataatggtttcttagacgtcaggtggcacttttcggggaaatgtgcgcggaacccctattt
gtttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattgaaaaaggaagagtatgag
tattcaacatttccgtgtcgcccttattccctttttgcggcattttgccttcctgtttttgctcacccagaaacgctggtgaaagtaaaagatg
ctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagttttcgccccgaagaacgt
tttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatac
actattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctg
ccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgcacaacatg
ggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgta
gcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcgga
taaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtggaagccgcg
gtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacg

aaatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagaccaagtttactcatatatactttagattgatttaaa
acttcatttttaatttaaaaggatctaggtgaagatccttttttgataatctcatgaccaaaatcccttaacgtgagttttcgttccactgagcgt
cagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctacc
agcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtcctt
ctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctg
ccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggggtt
cgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttccc
gaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacg
cctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaa
aaacgccagcaacgcggccttttacggttcctggccttttgctggccttttgctcacatgt

(c) Sequence of pAAV-CRISPR-Cas9 single-vector

cctgcaggcagctgcgcgctcgctcgctcactgaggccgcccgggcgtcgggcgacctttggtcgcccggcctcagtgagcgagc
gagcgcgcagagagggagtggccaactccatcactaggggttcctgcggcctCTAGAGGCTCCGGTGCCCGTC
AGTGGGCAGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCG
GCAATTGATCCGGTGCCTAGAGAAGGTGGCGCGGGGTAAACTGGGAAAGTGATG
TCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGC
AGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGGCc
ATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCAA
GCGGAACTACATCCTGGGCCTGGACATCGGCATCACCAGCGTGGGCTACGGCAT
CATCGACTACGAGACACGGGACGTGATCGATGCCGGCGTGCGGCTGTTCAAAGA
GGCCAACGTGGAAAACAACGAGGGCAGGCGGAGCAAGAGAGGCGCCAGAAGGC
TGAAGCGGCGGAGGCGGCATAGAATCCAGAGAGTGAAGAAGCTGCTGTTCGACT
ACAACCTGCTGACCGACCACAGCGAGCTGAGCGGCATCAACCCCTACGAGGCCA
GAGTGAAGGGCCTGAGCCAGAAGCTGAGCGAGGAAGAGTTCTCTGCCGCCCTGC
TGCACCTGGCCAAGAGAAGAGGCGTGCACAACGTGAACGAGGTGGAAGAGGAC
ACCGGCAACGAGCTGTCCACCAAAGAGCAGATCAGCCGGAACAGCAAGGCCCTG
GAAGAGAAATACGTGGCCGAACTGCAGCTGGAACGGCTGAAGAAAGACGGCGA
AGTGCGGGGCAGCATCAACAGATTCAAGACCAGCGACTACGTGAAAGAAGCCA

AACAGCTGCTGAAGGTGCAGAAGGCCTACCACCAGCTGGACCAGAGCTTCATCG
ACACCTACATCGACCTGCTGGAAACCCGGCGGACCTACTATGAGGGACCTGGCG
AGGGCAGCCCCTTCGGCTGGAAGGACATCAAAGAATGGTACGAGATGCTGATGG
GCCACTGCACCTACTTCCCCGAGGAACTGCGGAGCGTGAAGTACGCCTACAACG
CCGACCTGTACAACGCCCTGAACGACCTGAACAATCTCGTGATCACCAGGGACG
AGAACGAGAAGCTGGAATATTACGAGAAGTTCCAGATCATCGAGAACGTGTTCA
AGCAGAAGAAGAAGCCCACCCTGAAGCAGATCGCCAAAGAAATCCTCGTGAAC
GAAGAGGATATTAAGGGCTACAGAGTGACCAGCACCGGCAAGCCCGAGTTCACC
AACCTGAAGGTGTACCACGACATCAAGGACATTACCGCCCGGAAAGAGATTATT
GAGAACGCCGAGCTGCTGGATCAGATTGCCAAGATCCTGACCATCTACCAGAGC
AGCGAGGACATCCAGGAAGAACTGACCAATCTGAACTCCGAGCTGACCCAGGAA
GAGATCGAGCAGATCTCTAATCTGAAGGGCTATACCGGCACCCACAACCTGAGC
CTGAAGGCCATCAACCTGATCCTGGACGAGCTGTGGCACACCAACGACAACCAG
ATCGCTATCTTCAACCGGCTGAAGCTGGTGCCCAAGAAGGTGGACCTGTCCCAGC
AGAAAGAGATCCCCACCACCCTGGTGGACGACTTCATCCTGAGCCCCGTCGTGA
AGAGAAGCTTCATCCAGAGCATCAAAGTGATCAACGCCATCATCAAGAAGTACG
GCCTGCCCAACGACATCATTATCGAGCTGGCCCGCGAGAAGAACTCCAAGGACG
CCCAGAAAATGATCAACGAGATGCAGAAGCGGAACCGGCAGACCAACGAGCGG
ATCGAGGAAATCATCCGGACCACCGGCAAAGAGAACGCCAAGTACCTGATCGAG
AAGATCAAGCTGCACGACATGCAGGAAGGCAAGTGCCTGTACAGCCTGGAAGCC
ATCCCTCTGGAAGATCTGCTGAACAACCCCTTCAACTATGAGGTGGACCACATCA
TCCCCAGAAGCGTGTCCTTCGACAACAGCTTCAACAACAAGGTGCTCGTGAAGC
AGGAAGAAAACAGCAAGAAGGGCAACCGGACCCCATTCCAGTACCTGAGCAGC
AGCGACAGCAAGATCAGCTACGAAACCTTCAAGAAGCACATCCTGAATCTGGCC
AAGGGCAAGGGCAGAATCAGCAAGACCAAGAAAGAGTATCTGCTGGAAGAACG
GGACATCAACAGGTTCTCCGTGCAGAAAGACTTCATCAACCGGAACCTGGTGGA
TACCAGATACGCCACCAGAGGCCTGATGAACCTGCTGCGGAGCTACTTCAGAGT
GAACAACCTGGACGTGAAAGTGAAGTCCATCAATGGCGGCTTCACCAGCTTTCT
GCGGCGGAAGTGGAAGTTTAAGAAAGAGCGGAACAAGGGGTACAAGCACCACG
CCGAGGACGCCCTGATCATTGCCAACGCCGATTTCATCTTCAAAGAGTGGAAGA
AACTGGACAAGGCCAAAAAAGTGATGGAAAACCAGATGTTCGAGGAAAAGCAG
GCCGAGAGCATGCCCGAGATCGAAACCGAGCAGGAGTACAAAGAGATCTTCATC
ACCCCCCACCAGATCAAGCACATTAAGGACTTCAAGGACTACAAGTACAGCCAC
CGGGTGGACAAGAAGCCTAATAGAGAGCTGATTAACGACACCCTGTACTCCACC

CGGAAGGACGACAAGGGCAACACCCTGATCGTGAACAATCTGAACGGCCTGTAC
GACAAGGACAATGACAAGCTGAAAAAGCTGATCAACAAGAGCCCCGAAAAGCT
GCTGATGTACCACCACGACCCCCAGACCTACCAGAAACTGAAGCTGATTATGGA
ACAGTACGGCGACGAGAAGAATCCCCTGTACAAGTACTACGAGGAAACCGGGA
ACTACCTGACCAAGTACTCCAAAAAGGACAACGGCCCCGTGATCAAGAAGATTA
AGTATTACGGCAACAAACTGAACGCCCATCTGGACATCACCGACGACTACCCCA
ACAGCAGAAACAAGGTCGTGAAGCTGTCCCTGAAGCCCTACAGATTCGACGTGT
ACCTGGACAATGGCGTGTACAAGTTCGTGACCGTGAAGAATCTGGATGTGATCA
AAAAAGAAAACTACTACGAAGTGAATAGCAAGTGCTATGAGGAAGCTAAGAAG
CTGAAGAAGATCAGCAACCAGGCCGAGTTTATCGCCTCCTTCTACAACAACGATC
TGATCAAGATCAACGGCGAGCTGTATAGAGTGATCGGCGTGAACAACGACCTGC
TGAACCGGATCGAAGTGAACATGATCGACATCACCTACCGCGAGTACCTGGAAA
ACATGAACGACAAGAGGCCCCCCAGGATCATTAAGACAATCGCCTCCAAGACCC
AGAGCATTAAGAAGTACAGCACAGACATTCTGGGCAACCTGTATGAAGTGAAAT
CTAAGAAGCACCCTCAGATCATCAAAAAGGGCAAAAGGCCGGCGGCCACGAAA
AAGGCCGGCCAGGCAAAAAAGAAAAAGggatcctacccatacgatgttccagattacgcttacccatacga
tgttccagattacgcttacccatacgatgttccagattacgcttaagAATTCAATAAAAGATCTTTATTTTCATT
AGATCTGTGTGTTGGTTTTTTGTGTGCGGTACCTGAGGGCCTATTTCCCATGATTC
CTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGGAATTAATTT
GACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATT
TCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTA
CCGTAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACG
AAACACCGGAGACCACGGCAGGTCTCAGTTTTAGTACTCTGGAAACAGAATCTA
CTAAAACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGATTTTTG
CCTGAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGT
TAGAGAGATAATTGGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAA
TACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGT
TTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGC
TTTATATATCTTGTGGAAAGGACGAAACACCGGGTCTTCGGATCCATGCGAAGAC
CTGTTTTAGTACTCTGGAAACAGAATCTACTAAAACAAGGCAAAATGCCGTGTTT
ATCTCGTCAACTTGTTGGCGAGATTTTTGCggccgcaggaaccctagtgatggagttggccactccctct
ctgcgcgctcgctcgctcactgaggccgggcgaccaaaggtcgcccgacgcccgggctttgcccgggcggcctcagtgagcgag
cgagcgcgcagctgcctgcaggggcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgcatacgtcaaagcaa
ccatagtacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccct

agcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggtt
ccgatttagtgctttacggcacctcgaccccaaaaaacttgatttgggtgatggttcacgtagtgggccatcgccctgatagacggtttttt
cgcccttttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcgggctattcttttgatt
tataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttac
aattttatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagccccgacacccgccaacacccgctgacgcgccct
gacgggcttgtctgctcccggcatccgcttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatca
ccgaaacgcgcgagacgaaagggcctcgtgatacgcctatttttataggttaatgtcatgataataatggtttcttagacgtcaggtggc
actttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaa
tgcttcaataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattccctttttgcggcattttgccttcctgtttttgc
tcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggt
aagatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgac
gccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacgg
atggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggagg
accgaaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccatac
caaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcc
cggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctga
taaatctggagccggtgagcgtggaagccgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctac
acgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcag
accaagtttactcatatatactttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaagatccttttgataatctcatgaccaa
aatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaat
ctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactgg
cttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacata
cctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccgga
taaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagataccta
cagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacagg
agagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttt
gtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcctggccttttgctggccttttgctca
catgt

(d) Sequence of pC901

cctgcaggcagctgcgcgctcgctcgctcactgaggccgcccgggcgtcgggcgacctttggtcgcccggcctcagtgagcgagc
gagcgcgcagagagggagtggccaactccatcactaggggttcctgcggcctCTAGAGGCTCCGGTGCCCGTC
AGTGGGCAGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCG
GCAATTGATCCGGTGCCTAGAGAAGGTGGCGCGGGGTAAACTGGGAAAGTGATG
TCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGC
AGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGGCc
ATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCAA
GCGGAACTACATCCTGGGCCTGGACATCGGCATCACCAGCGTGGGCTACGGCAT
CATCGACTACGAGACACGGGACGTGATCGATGCCGGCGTGCGGCTGTTCAAAGA
GGCCAACGTGGAAAACAACGAGGGCAGGCGGAGCAAGAGAGGCGCCAGAAGGC
TGAAGCGGCGGAGGCGGCATAGAATCCAGAGAGTGAAGAAGCTGCTGTTCGACT
ACAACCTGCTGACCGACCACAGCGAGCTGAGCGGCATCAACCCCTACGAGGCCA
GAGTGAAGGGCCTGAGCCAGAAGCTGAGCGAGGAAGAGTTCTCTGCCGCCCTGC
TGCACCTGGCCAAGAGAAGAGGCGTGCACAACGTGAACGAGGTGGAAGAGGAC
ACCGGCAACGAGCTGTCCACCAAAGAGCAGATCAGCCGGAACAGCAAGGCCCTG
GAAGAGAAATACGTGGCCGAACTGCAGCTGGAACGGCTGAAGAAAGACGGCGA
AGTGCGGGGCAGCATCAACAGATTCAAGACCAGCGACTACGTGAAAGAAGCCA
AACAGCTGCTGAAGGTGCAGAAGGCCTACCACCAGCTGGACCAGAGCTTCATCG
ACACCTACATCGACCTGCTGGAAACCCGGCGGACCTACTATGAGGGACCTGGCG
AGGGCAGCCCCTTCGGCTGGAAGGACATCAAAGAATGGTACGAGATGCTGATGG
GCCACTGCACCTACTTCCCCGAGGAACTGCGGAGCGTGAAGTACGCCTACAACG
CCGACCTGTACAACGCCCTGAACGACCTGAACAATCTCGTGATCACCAGGGACG
AGAACGAGAAGCTGGAATATTACGAGAAGTTCCAGATCATCGAGAACGTGTTCA
AGCAGAAGAAGAAGCCCACCCTGAAGCAGATCGCCAAAGAAATCCTCGTGAAC
GAAGAGGATATTAAGGGCTACAGAGTGACCAGCACCGGCAAGCCCGAGTTCACC
AACCTGAAGGTGTACCACGACATCAAGGACATTACCGCCCGGAAAGAGATTATT
GAGAACGCCGAGCTGCTGGATCAGATTGCCAAGATCCTGACCATCTACCAGAGC
AGCGAGGACATCCAGGAAGAACTGACCAATCTGAACTCCGAGCTGACCCAGGAA
GAGATCGAGCAGATCTCTAATCTGAAGGGCTATACCGGCACCCACAACCTGAGC
CTGAAGGCCATCAACCTGATCCTGGACGAGCTGTGGCACACCAACGACAACCAG
ATCGCTATCTTCAACCGGCTGAAGCTGGTGCCCAAGAAGGTGGACCTGTCCCAGC

AGAAAGAGATCCCCACCACCCTGGTGGACGACTTCATCCTGAGCCCCGTCGTGA
AGAGAAGCTTCATCCAGAGCATCAAAGTGATCAACGCCATCATCAAGAAGTACG
GCCTGCCCAACGACATCATTATCGAGCTGGCCCGCGAGAAGAACTCCAAGGACG
CCCAGAAAATGATCAACGAGATGCAGAAGCGGAACCGGCAGACCAACGAGCGG
ATCGAGGAAATCATCCGGACCACCGGCAAAGAGAACGCCAAGTACCTGATCGAG
AAGATCAAGCTGCACGACATGCAGGAAGGCAAGTGCCTGTACAGCCTGGAAGCC
ATCCCTCTGGAAGATCTGCTGAACAACCCCTTCAACTATGAGGTGGACCACATCA
TCCCCAGAAGCGTGTCCTTCGACAACAGCTTCAACAACAAGGTGCTCGTGAAGC
AGGAAGAAAACAGCAAGAAGGGCAACCGGACCCCATTCCAGTACCTGAGCAGC
AGCGACAGCAAGATCAGCTACGAAACCTTCAAGAAGCACATCCTGAATCTGGCC
AAGGGCAAGGGCAGAATCAGCAAGACCAAGAAAGAGTATCTGCTGGAAGAACG
GGACATCAACAGGTTCTCCGTGCAGAAAGACTTCATCAACCGGAACCTGGTGGA
TACCAGATACGCCACCAGAGGCCTGATGAACCTGCTGCGGAGCTACTTCAGAGT
GAACAACCTGGACGTGAAAGTGAAGTCCATCAATGGCGGCTTCACCAGCTTTCT
GCGGCGGAAGTGGAAGTTTAAGAAAGAGCGGAACAAGGGGTACAAGCACCACG
CCGAGGACGCCCTGATCATTGCCAACGCCGATTTCATCTTCAAAGAGTGGAAGA
AACTGGACAAGGCCAAAAAAGTGATGGAAAACCAGATGTTCGAGGAAAAGCAG
GCCGAGAGCATGCCCGAGATCGAAACCGAGCAGGAGTACAAAGAGATCTTCATC
ACCCCCCACCAGATCAAGCACATTAAGGACTTCAAGGACTACAAGTACAGCCAC
CGGGTGGACAAGAAGCCTAATAGAGAGCTGATTAACGACACCCTGTACTCCACC
CGGAAGGACGACAAGGGCAACACCCTGATCGTGAACAATCTGAACGGCCTGTAC
GACAAGGACAATGACAAGCTGAAAAAGCTGATCAACAAGAGCCCCGAAAAGCT
GCTGATGTACCACCACGACCCCCAGACCTACCAGAAACTGAAGCTGATTATGGA
ACAGTACGGCGACGAGAAGAATCCCCTGTACAAGTACTACGAGGAAACCGGGA
ACTACCTGACCAAGTACTCCAAAAAGGACAACGGCCCCGTGATCAAGAAGATTA
AGTATTACGGCAACAAACTGAACGCCCATCTGGACATCACCGACGACTACCCCA
ACAGCAGAAACAAGGTCGTGAAGCTGTCCCTGAAGCCCTACAGATTCGACGTGT
ACCTGGACAATGGCGTGTACAAGTTCGTGACCGTGAAGAATCTGGATGTGATCA
AAAAAGAAAACTACTACGAAGTGAATAGCAAGTGCTATGAGGAAGCTAAGAAG
CTGAAGAAGATCAGCAACCAGGCCGAGTTTATCGCCTCCTTCTACAACAACGATC
TGATCAAGATCAACGGCGAGCTGTATAGAGTGATCGGCGTGAACAACGACCTGC
TGAACCGGATCGAAGTGAACATGATCGACATCACCTACCGCGAGTACCTGGAAA
ACATGAACGACAAGAGGCCCCCCAGGATCATTAAGACAATCGCCTCCAAGACCC
AGAGCATTAAGAAGTACAGCACAGACATTCTGGGCAACCTGTATGAAGTGAAAT

CTAAGAAGCACCCTCAGATCATCAAAAAGGGCAAAAGGCCGGCGGCCACGAAA
AAGGCCGGCCAGGCAAAAAAGAAAAAGggatcctacccatacgatgttccagattacgcttacccatacga
tgttccagattacgcttacccatacgatgttccagattacgcttaagAATTCAATAAAAGATCTTTATTTTCATT
AGATCTGTGTGTTGGTTTTTTGTGTGCGGTACCTGAGGGCCTATTTCCCATGATTC
CTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGGAATTAATTT
GACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATT
TCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTA
CCGTAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACG
AAACACCGCTTGCTCTCACAGTACTCGCTGTTTTAGTACTCTGGAAACAGAATCT
ACTAAAACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGATTTTT
GCCTGAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCT
GTTAGAGAGATAATTGGAATTAATTTGACTGTAAACACAAAGATATTAGTACAA
AATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTAT
GTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTG
GCTTTATATATCTTGTGGAAAGGACGAAACACCGGCGCAGGCGGTGGCGAGTGG
GTTTTAGTACTCTGGAAACAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTAT
CTCGTCAACTTGTTGGCGAGATTTTTGCggccgcaggaacccctagtgatggagttggccactccctctctg
cgcgctcgctcgctcactgaggccgggcgaccaaaggtcgcccgacgcccgggctttgcccgggcggcctcagtgagcgagcga
gcgcgcagctgcctgcaggggcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgcatacgtcaaagcaacca
tagtacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagc
gcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccg
atttagtgctttacggcacctcgaccccaaaaaacttgatttgggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgc
cctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcgggctattcttttgatttata
agggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaattt
tatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagccccgacacccgccaacacccgctgacgcgccctgacg
ggcttgtctgctcccggcatccgcttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccga
aacgcgcgagacgaaagggcctcgtgatacgcctatttttataggttaatgtcatgataataatggtttcttagacgtcaggtggcacttttt
cggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatgctt
caataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattccctttttttgcggcattttgccttcctgtttttgctcac
ccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaag
atccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgcc
gggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatg
gcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggacc
gaaggagctaaccgctttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaa

acgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccg
gcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgata
aatctggagccggtgagcgtggaagccgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacac
gacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagac
caagtttactcatatatactttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaagatccttttttgataatctcatgaccaaaa
tcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatct
gctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggct
tcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacc
tcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggata
aggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctaca
gcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggag
agcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgt
gatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcctggccttttgctggccttttgctcac
atgt

(e) Sequence of pC902

cctgcaggcagctgcgcgctcgctcgctcactgaggccgcccgggcgtcgggcgacctttggtcgcccggcctcagtgagcgagc
gagcgcgcagagagggagtggccaactccatcactaggggttcctgcggcctCTAGAGGCTCCGGTGCCCGTC
AGTGGGCAGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCG
GCAATTGATCCGGTGCCTAGAGAAGGTGGCGCGGGGTAAACTGGGAAAGTGATG
TCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGC
AGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGGCc
ATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCAA
GCGGAACTACATCCTGGGCCTGGACATCGGCATCACCAGCGTGGGCTACGGCAT
CATCGACTACGAGACACGGGACGTGATCGATGCCGGCGTGCGGCTGTTCAAAGA
GGCCAACGTGGAAAACAACGAGGGCAGGCGGAGCAAGAGAGGCGCCAGAAGGC
TGAAGCGGCGGAGGCGGCATAGAATCCAGAGAGTGAAGAAGCTGCTGTTCGACT
ACAACCTGCTGACCGACCACAGCGAGCTGAGCGGCATCAACCCCTACGAGGCCA
GAGTGAAGGGCCTGAGCCAGAAGCTGAGCGAGGAAGAGTTCTCTGCCGCCCTGC

TGCACCTGGCCAAGAGAAGAGGCGTGCACAACGTGAACGAGGTGGAAGAGGAC
ACCGGCAACGAGCTGTCCACCAAAGAGCAGATCAGCCGGAACAGCAAGGCCCTG
GAAGAGAAATACGTGGCCGAACTGCAGCTGGAACGGCTGAAGAAAGACGGCGA
AGTGCGGGGCAGCATCAACAGATTCAAGACCAGCGACTACGTGAAAGAAGCCA
AACAGCTGCTGAAGGTGCAGAAGGCCTACCACCAGCTGGACCAGAGCTTCATCG
ACACCTACATCGACCTGCTGGAAACCCGGCGGACCTACTATGAGGGACCTGGCG
AGGGCAGCCCCTTCGGCTGGAAGGACATCAAAGAATGGTACGAGATGCTGATGG
GCCACTGCACCTACTTCCCCGAGGAACTGCGGAGCGTGAAGTACGCCTACAACG
CCGACCTGTACAACGCCCTGAACGACCTGAACAATCTCGTGATCACCAGGGACG
AGAACGAGAAGCTGGAATATTACGAGAAGTTCCAGATCATCGAGAACGTGTTCA
AGCAGAAGAAGAAGCCCACCCTGAAGCAGATCGCCAAAGAAATCCTCGTGAAC
GAAGAGGATATTAAGGGCTACAGAGTGACCAGCACCGGCAAGCCCGAGTTCACC
AACCTGAAGGTGTACCACGACATCAAGGACATTACCGCCCGGAAAGAGATTATT
GAGAACGCCGAGCTGCTGGATCAGATTGCCAAGATCCTGACCATCTACCAGAGC
AGCGAGGACATCCAGGAAGAACTGACCAATCTGAACTCCGAGCTGACCCAGGAA
GAGATCGAGCAGATCTCTAATCTGAAGGGCTATACCGGCACCCACAACCTGAGC
CTGAAGGCCATCAACCTGATCCTGGACGAGCTGTGGCACACCAACGACAACCAG
ATCGCTATCTTCAACCGGCTGAAGCTGGTGCCCAAGAAGGTGGACCTGTCCCAGC
AGAAAGAGATCCCCACCACCCTGGTGGACGACTTCATCCTGAGCCCCGTCGTGA
AGAGAAGCTTCATCCAGAGCATCAAAGTGATCAACGCCATCATCAAGAAGTACG
GCCTGCCCAACGACATCATTATCGAGCTGGCCCGCGAGAAGAACTCCAAGGACG
CCCAGAAAATGATCAACGAGATGCAGAAGCGGAACCGGCAGACCAACGAGCGG
ATCGAGGAAATCATCCGGACCACCGGCAAAGAGAACGCCAAGTACCTGATCGAG
AAGATCAAGCTGCACGACATGCAGGAAGGCAAGTGCCTGTACAGCCTGGAAGCC
ATCCCTCTGGAAGATCTGCTGAACAACCCCTTCAACTATGAGGTGGACCACATCA
TCCCCAGAAGCGTGTCCTTCGACAACAGCTTCAACAACAAGGTGCTCGTGAAGC
AGGAAGAAAACAGCAAGAAGGGCAACCGGACCCCATTCCAGTACCTGAGCAGC
AGCGACAGCAAGATCAGCTACGAAACCTTCAAGAAGCACATCCTGAATCTGGCC
AAGGGCAAGGGCAGAATCAGCAAGACCAAGAAAGAGTATCTGCTGGAAGAACG
GGACATCAACAGGTTCTCCGTGCAGAAAGACTTCATCAACCGGAACCTGGTGGA
TACCAGATACGCCACCAGAGGCCTGATGAACCTGCTGCGGAGCTACTTCAGAGT
GAACAACCTGGACGTGAAAGTGAAGTCCATCAATGGCGGCTTCACCAGCTTTCT
GCGGCGGAAGTGGAAGTTTAAGAAAGAGCGGAACAAGGGGTACAAGCACCACG
CCGAGGACGCCCTGATCATTGCCAACGCCGATTTCATCTTCAAAGAGTGGAAGA

AACTGGACAAGGCCAAAAAAGTGATGGAAAACCAGATGTTCGAGGAAAAGCAG
GCCGAGAGCATGCCCGAGATCGAAACCGAGCAGGAGTACAAAGAGATCTTCATC
ACCCCCCACCAGATCAAGCACATTAAGGACTTCAAGGACTACAAGTACAGCCAC
CGGGTGGACAAGAAGCCTAATAGAGAGCTGATTAACGACACCCTGTACTCCACC
CGGAAGGACGACAAGGGCAACACCCTGATCGTGAACAATCTGAACGGCCTGTAC
GACAAGGACAATGACAAGCTGAAAAAGCTGATCAACAAGAGCCCCGAAAAGCT
GCTGATGTACCACCACGACCCCCAGACCTACCAGAAACTGAAGCTGATTATGGA
ACAGTACGGCGACGAGAAGAATCCCCTGTACAAGTACTACGAGGAAACCGGGA
ACTACCTGACCAAGTACTCCAAAAAGGACAACGGCCCCGTGATCAAGAAGATTA
AGTATTACGGCAACAAACTGAACGCCCATCTGGACATCACCGACGACTACCCCA
ACAGCAGAAACAAGGTCGTGAAGCTGTCCCTGAAGCCCTACAGATTCGACGTGT
ACCTGGACAATGGCGTGTACAAGTTCGTGACCGTGAAGAATCTGGATGTGATCA
AAAAAGAAAACTACTACGAAGTGAATAGCAAGTGCTATGAGGAAGCTAAGAAG
CTGAAGAAGATCAGCAACCAGGCCGAGTTTATCGCCTCCTTCTACAACAACGATC
TGATCAAGATCAACGGCGAGCTGTATAGAGTGATCGGCGTGAACAACGACCTGC
TGAACCGGATCGAAGTGAACATGATCGACATCACCTACCGCGAGTACCTGGAAA
ACATGAACGACAAGAGGCCCCCCAGGATCATTAAGACAATCGCCTCCAAGACCC
AGAGCATTAAGAAGTACAGCACAGACATTCTGGGCAACCTGTATGAAGTGAAAT
CTAAGAAGCACCCTCAGATCATCAAAAAGGGCAAAAGGCCGGCGGCCACGAAA
AAGGCCGGCCAGGCAAAAAAGAAAAAGggatcctacccatacgatgttccagattacgcttacccatacga
tgttccagattacgcttacccatacgatgttccagattacgcttaagAATTCAATAAAAGATCTTTATTTTCATT
AGATCTGTGTGTTGGTTTTTTGTGTGCGGTACCTGAGGGCCTATTTCCCATGATTC
CTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGGAATTAATTT
GACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATT
TCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTA
CCGTAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACG
AAACACCGATCAGTTCGACGCGATACGGAGTTTTAGTACTCTGGAAACAGAATC
TACTAAAACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGATTTT
TGCCTGAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCT
GTTAGAGAGATAATTGGAATTAATTTGACTGTAAACACAAAGATATTAGTACAA
AATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTAT
GTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTG
GCTTTATATATCTTGTGGAAAGGACGAAACACCGCGATCCGAATAGCTCGATGC
GTTTTAGTACTCTGGAAACAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTAT

CTCGTCAACTTGTTGGCGAGATTTTTGCggccgcaggaacccctagtgatggagttggccactccctctctg
cgcgctcgctcgctcactgaggccgggcgaccaaaggtcgcccgacgcccgggctttgcccgggcggcctcagtgagcgagcga
gcgcgcagctgcctgcaggggcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgcatacgtcaaagcaacca
tagtacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagc
gcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccg
atttagtgctttacggcacctcgaccccaaaaaacttgatttgggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgc
cctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcgggctattcttttgatttata
agggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaattt
tatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagccccgacacccgccaacacccgctgacgcgccctgacg
ggcttgtctgctcccggcatccgcttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccga
aacgcgcgagacgaaagggcctcgtgatacgcctatttttataggttaatgtcatgataataatggtttcttagacgtcaggtggcacttttt
cggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatgctt
caataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattccctttttgcggcattttgccttcctgtttttgctcac
ccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaag
atccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgcc
gggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatg
gcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggacc
gaaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaa
acgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccg
gcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgata
aatctggagccggtgagcgtggaagccgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacac
gacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagac
caagtttactcatatatactttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaagatccttttgataatctcatgaccaaaa
tcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatct
gctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggct
tcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacc
tcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggata
aggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctaca
gcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggag
agcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgt
gatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggcctttttacggttcctggccttttgctggccttttgctcac
atgt

(f) Sequence of pC904

cctgcaggcagctgcgcgctcgctcgctcactgaggccgcccgggcgtcgggcgacctttggtcgcccggcctcagtgagcgagc
gagcgcgcagagagggagtggccaactccatcactaggggttcctgcggcctCTAGAGGCTCCGGTGCCCGTC
AGTGGGCAGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCG
GCAATTGATCCGGTGCCTAGAGAAGGTGGCGCGGGGTAAACTGGGAAAGTGATG
TCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGC
AGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGGCc
ATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCAA
GCGGAACTACATCCTGGGCCTGGACATCGGCATCACCAGCGTGGGCTACGGCAT
CATCGACTACGAGACACGGGACGTGATCGATGCCGGCGTGCGGCTGTTCAAAGA
GGCCAACGTGGAAAACAACGAGGGCAGGCGGAGCAAGAGAGGCGCCAGAAGGC
TGAAGCGGCGGAGGCGGCATAGAATCCAGAGAGTGAAGAAGCTGCTGTTCGACT
ACAACCTGCTGACCGACCACAGCGAGCTGAGCGGCATCAACCCCTACGAGGCCA
GAGTGAAGGGCCTGAGCCAGAAGCTGAGCGAGGAAGAGTTCTCTGCCGCCCTGC
TGCACCTGGCCAAGAGAAGAGGCGTGCACAACGTGAACGAGGTGGAAGAGGAC
ACCGGCAACGAGCTGTCCACCAAAGAGCAGATCAGCCGGAACAGCAAGGCCCTG
GAAGAGAAATACGTGGCCGAACTGCAGCTGGAACGGCTGAAGAAAGACGGCGA
AGTGCGGGGCAGCATCAACAGATTCAAGACCAGCGACTACGTGAAAGAAGCCA
AACAGCTGCTGAAGGTGCAGAAGGCCTACCACCAGCTGGACCAGAGCTTCATCG
ACACCTACATCGACCTGCTGGAAACCCGGCGGACCTACTATGAGGGACCTGGCG
AGGGCAGCCCCTTCGGCTGGAAGGACATCAAAGAATGGTACGAGATGCTGATGG
GCCACTGCACCTACTTCCCCGAGGAACTGCGGAGCGTGAAGTACGCCTACAACG
CCGACCTGTACAACGCCCTGAACGACCTGAACAATCTCGTGATCACCAGGGACG
AGAACGAGAAGCTGGAATATTACGAGAAGTTCCAGATCATCGAGAACGTGTTCA
AGCAGAAGAAGAAGCCCACCCTGAAGCAGATCGCCAAAGAAATCCTCGTGAAC
GAAGAGGATATTAAGGGCTACAGAGTGACCAGCACCGGCAAGCCCGAGTTCACC
AACCTGAAGGTGTACCACGACATCAAGGACATTACCGCCCGGAAAGAGATTATT
GAGAACGCCGAGCTGCTGGATCAGATTGCCAAGATCCTGACCATCTACCAGAGC
AGCGAGGACATCCAGGAAGAACTGACCAATCTGAACTCCGAGCTGACCCAGGAA
GAGATCGAGCAGATCTCTAATCTGAAGGGCTATACCGGCACCCACAACCTGAGC

CTGAAGGCCATCAACCTGATCCTGGACGAGCTGTGGCACACCAACGACAACCAG
ATCGCTATCTTCAACCGGCTGAAGCTGGTGCCCAAGAAGGTGGACCTGTCCCAGC
AGAAAGAGATCCCCACCACCCTGGTGGACGACTTCATCCTGAGCCCCGTCGTGA
AGAGAAGCTTCATCCAGAGCATCAAAGTGATCAACGCCATCATCAAGAAGTACG
GCCTGCCCAACGACATCATTATCGAGCTGGCCCGCGAGAAGAACTCCAAGGACG
CCCAGAAAATGATCAACGAGATGCAGAAGCGGAACCGGCAGACCAACGAGCGG
ATCGAGGAAATCATCCGGACCACCGGCAAAGAGAACGCCAAGTACCTGATCGAG
AAGATCAAGCTGCACGACATGCAGGAAGGCAAGTGCCTGTACAGCCTGGAAGCC
ATCCCTCTGGAAGATCTGCTGAACAACCCCTTCAACTATGAGGTGGACCACATCA
TCCCCAGAAGCGTGTCCTTCGACAACAGCTTCAACAACAAGGTGCTCGTGAAGC
AGGAAGAAAACAGCAAGAAGGGCAACCGGACCCCATTCCAGTACCTGAGCAGC
AGCGACAGCAAGATCAGCTACGAAACCTTCAAGAAGCACATCCTGAATCTGGCC
AAGGGCAAGGGCAGAATCAGCAAGACCAAGAAAGAGTATCTGCTGGAAGAACG
GGACATCAACAGGTTCTCCGTGCAGAAAGACTTCATCAACCGGAACCTGGTGGA
TACCAGATACGCCACCAGAGGCCTGATGAACCTGCTGCGGAGCTACTTCAGAGT
GAACAACCTGGACGTGAAAGTGAAGTCCATCAATGGCGGCTTCACCAGCTTTCT
GCGGCGGAAGTGGAAGTTTAAGAAAGAGCGGAACAAGGGGTACAAGCACCACG
CCGAGGACGCCCTGATCATTGCCAACGCCGATTTCATCTTCAAAGAGTGGAAGA
AACTGGACAAGGCCAAAAAAGTGATGGAAAACCAGATGTTCGAGGAAAAGCAG
GCCGAGAGCATGCCCGAGATCGAAACCGAGCAGGAGTACAAAGAGATCTTCATC
ACCCCCCACCAGATCAAGCACATTAAGGACTTCAAGGACTACAAGTACAGCCAC
CGGGTGGACAAGAAGCCTAATAGAGAGCTGATTAACGACACCCTGTACTCCACC
CGGAAGGACGACAAGGGCAACACCCTGATCGTGAACAATCTGAACGGCCTGTAC
GACAAGGACAATGACAAGCTGAAAAAGCTGATCAACAAGAGCCCCGAAAAGCT
GCTGATGTACCACCACGACCCCCAGACCTACCAGAAACTGAAGCTGATTATGGA
ACAGTACGGCGACGAGAAGAATCCCCTGTACAAGTACTACGAGGAAACCGGGA
ACTACCTGACCAAGTACTCCAAAAAGGACAACGGCCCCGTGATCAAGAAGATTA
AGTATTACGGCAACAAACTGAACGCCCATCTGGACATCACCGACGACTACCCCA
ACAGCAGAAACAAGGTCGTGAAGCTGTCCCTGAAGCCCTACAGATTCGACGTGT
ACCTGGACAATGGCGTGTACAAGTTCGTGACCGTGAAGAATCTGGATGTGATCA
AAAAAGAAAACTACTACGAAGTGAATAGCAAGTGCTATGAGGAAGCTAAGAAG
CTGAAGAAGATCAGCAACCAGGCCGAGTTTATCGCCTCCTTCTACAACAACGATC
TGATCAAGATCAACGGCGAGCTGTATAGAGTGATCGGCGTGAACAACGACCTGC
TGAACCGGATCGAAGTGAACATGATCGACATCACCTACCGCGAGTACCTGGAAA

ACATGAACGACAAGAGGCCCCCCAGGATCATTAAGACAATCGCCTCCAAGACCC
AGAGCATTAAGAAGTACAGCACAGACATTCTGGGCAACCTGTATGAAGTGAAAT
CTAAGAAGCACCCTCAGATCATCAAAAAGGGCAAAAGGCCGGCGGCCACGAAA
AAGGCCGGCCAGGCAAAAAAGAAAAAGggatcctacccatacgatgttccagattacgcttacccatacga
tgttccagattacgcttacccatacgatgttccagattacgcttaagaattcctagagctcgctgatcagcctcgactgtgccttctagttgc
cagccatctgttgtttgcccctcccccgtgccttccttgaccctggaaggtgccactcccactgtcctttcctaataaaatgaggaaattgc
atcgcattgtctgagtaggtgtcattctattctgggggggtggggtggggcaggacagcaaggggaggattgggaagagaatagca
ggcatgctggggaggtacCTGAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGA
TACAAGGCTGTTAGAGAGATAATTGGAATTAATTTGACTGTAAACACAAAGATA
TTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTT
TAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTT
CGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGCTTGCTCTCAC
AGTACTCGCTGTTTTAGTACTCTGGAAACAGAATCTACTAAAACAAGGCAAAAT
GCCGTGTTTATCTCGTCAACTTGTTGGCGAGATTTTTGCCTGAGGGCCTATTTCCC
ATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGGAA
TTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGT
AATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCAT
ATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAA
AGGACGAAACACCGCTGTAGCAAGCTCTGGAACTCGTTTTAGTACTCTGGAAAC
AGAATCTACTAAAACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGA
GATTTTTGCggccgcaggaaccctagtgatggagttggccactccctctctgcgcgctcgctcgctcactgaggccgggcg
accaaaggtcgcccgacgcccgggctttgcccgggcggcctcagtgagcgagcgagcgcgcagctgcctgcaggggcgcctgat
gcggtattttctccttacgcatctgtgcggtatttcacaccgcatacgtcaaagcaaccatagtacgcgccctgtagcggcgcattaagc
gcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctc
gccacgttcgccggctttccccgtcaagctctaaatcgggggctcccttagggttccgatttagtgctttacggcacctcgaccccaaa
aaacttgatttgggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagt
ggactcttgttccaaactggaacaacactcaaccctatctcgggctattcttttgatttataagggattttgccgatttcggcctattggttaa
aaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaattttatggtgcactctcagtacaatctgctctgat
gccgcatagttaagccagccccgacacccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggcatccgcttacag
acaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgagacgaaagggcctcgtgat
acgcctatttttataggttaatgtcatgataataatggtttcttagacgtcaggtggcacttttcggggaaatgtgcgcggaacccctatttgt
ttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattgaaaaaggaagagtatgagta
ttcaacatttccgtgtcgcccttattccctttttgcggcattttgccttcctgtttttgctcacccagaaacgctggtgaaagtaaaagatgct
gaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagttttcgccccgaagaacgttt

tccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcataca
ctattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgc
cataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgcacaacatgg
gggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtag
caatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcggat
aaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtggaagccgcg
gtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacg
aaatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagaccaagtttactcatatatactttagattgatttaaa
acttcatttttaatttaaaaggatctaggtgaagatccttttgataatctcatgaccaaaatcccttaacgtgagttttcgttccactgagcgt
cagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctacc
agcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtcctt
ctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctg
ccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggggtt
cgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttccc
gaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacg
cctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaa
aaacgccagcaacgcggccttttacggttcctggccttttgctggccttttgctcacatgt

Appendix 2. Sanger sequencing of DNA bands.

(a) Sequencing of unedited band

NNNNNNNNNNGGNNNNNGATTTNNCGNCCGCGAATTCGCCCTTACCNGTCGCT
AGAGGCGAAAGCCCGACACCCAGCTTCGGTCAGAGAAATGAGAGGGAAAGTAA
AAATGCGTCGAGCTCTGAGGAGAGCCCCCGCTTCTACCCGCGCCTCTTCCCGGCA
GCCGAACCCCAAACAGCCACCCGCCAGGATGCCGCCTCCTCACTCACCCACTCGC
CACCGCCTGCGCCTCCGCCGCCGCGGGCGCAGGCACCGCAACCGCAGCCCCGCC
CCGGGCCCGCCCCCGGGCCCGCCCCGACCACGCCCCGGCCCCGGCCCCTAGCGC
GCGACTCCTGAGTTCCAGAGCTTGCTACAGGCTGCGGTTGTTTCCCTCCTTGTTTT

CTTCTGGTTAATCTTTATCAGGTCTTTTCTTGTTCACCCTCAGCGAGTACTGTGAG
AGCAAGTAGTGGGGAGAGAGGGTGGGAAAAACAAAAACACACACCTCCTAAAC
CCACACCTGCTCTTGCTTGACCCCGCCCCCAAAGGAGAAGCAACCGGGCAGCAG
GGACGGCTGACACACCAAGCGTCATCTTTTACGTGGGCGGAACTTGTCGCTGTTT
GACGCACCTCTCTTTCCTAGCGGGACACCGTANGTTACGTCTGTCTGTTTTCTATG
TGCGATGACNAANGNCGAATTCGTTTAAACCTGCANGACTAGTCCCTTTANTGAG
GGTTAATTCTGANCTTGGCNTAATCATGGTCATANCTGTTNCCTGTGTGAAATNN
TTANCCNCNNNNNTTCCACNCAACANACNAGCCGGAANGCATAANNNGNNAAN
NNNNNGGNGNNNANNGAGNNNNN

(b) Sequencing of T4B/T5B edited band

NNNNNNNNNNNNGNNNNTGNTTTNNCGNCCGCGAATTCGCCCTTCGTCATCGC
ACATAGAAAACAGACAGACGTAACCTACGGTGTCCCGCTAGGAAAGAGAGGTGC
GTCAAACAGCGACAAGTTCCGCCCACGTAAAAGATGACGCTTGGTGTGTCAGCC
GTCCCTGCTGCCCGGTTGCTTCTCTTTTGGGGGCGGGGTCTAGCAAGAGCATGGG
TGAGTGAGGAGGCGGCATCCTGGCGGGTGGCTGTTTGGGGTTCGGCTGCCGGGA
AGAGGCGCGGGTAGAAGCGGGGGCTCTCCTCAGAGCTCGACGCATTTTTACTTTC
CCTCTCATTTCTCTGACCGAAGCTGGGTGTCGGGCTTTCGCCTCTAGCGACTGGT
AAGGGCGAATTCGTTTAAACCTGCAGGACTAGTCCCTTTAGTGAGGGTTAATTCT
GAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGNGTGAAATTGTTATCCGCTC
ACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCC
TAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCANT
CGGGAANCNTNNNNTGCCANCTGCATTANNNNNTCNGCCAANGCGCGGNNNNN
NNCGNTNNNNNNNNNNNGCNNNCTNNNNNTNNNTCNNNCNNTNNNTNNNNTNN
NNNNNNNNNNNN

(c) Sequencing of T4I/T5B edited band

NNNNNNNNNNNNGGNNANNGNNNTNNCGNCCGCGAATTCGCCCTTCGTCATCGC
ACATAGAAAACAGACAGACGTAACCTACGGTGTCCCGCTAGGAAAGAGAGGTGC
GTCAAACAGCGACAAGTTCCGCCCACGTAAAAGATGACGCTTGGTGTGTCAGCC
GTCCCTGCTGCCCGGTTGCTTCTCTTTTGGGGGCGGGGTCTAGCAAGAGCAGGTG
TGGGTTTAGGAGGTGTGTGTTTTTGTTTTTCCCACCCTCTCTCCCCACTACTTGCT
CTCACAGTACTCTGGGTGAGTGAGGAGGCGGCATCCTGGCGGGTGGCTGTTTGG
GGTTCGGCTGCCGGGAAGAGGCGCGGGTAGAAGCGGGGGCTCTCCTCAGAGCTC
GACGCATTTTTACTTTCCCTCTCATTTCTCTGACCGAAGCTGGGTGTCGGGCTTTC
GCCTCTAGCGACTGGTAAGGGCGAATTCGTTTAAACCTGCAGGACTAGTCCCTTT
AGTGAGGGTTAATTCTGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTG
AAATTGTTATCCGCTCACAATTCCACACNANATACGAGCCGGANGCATAAAGTG
TAAAGCCTGNGGTGNNNNNNGAGTGAGCTAACTNNNNNTAATTGCGTTGCGCTC
ACTGCCCGCTTTCNNGTCNGGAANNCTGTCNNGNCNNNNTGCATNNNNNNNNNN
NNNANNCNNGGNNNNNN

Appendix 3. Sanger sequencing of C9-500 BAC amplified G4C2 flanking sequences.

(a) Sequencing of G4C2 upstream region

NNNNNNNNGGNNATTGATTTAGCGGCCGCGAATTCGCCCTTCGACTCCTGAGT
TCCAGAGCTTGCTACAGGCTGCGGTTGTTTCCCTCCTTGTTTTCTTCTGGTTAATC
TTTATCAGGTCTTTTCTTGTTCACCCTCAGCGAGTACTGTGAGAGCAAGTAGTGG
GGAGAGAGGGTGGGAAAAACAAAAACACACACCTCCTAAACCCACACCAAGGG
CGAATTCGTTTAAACCTGCAGGACTAGTCCCTTTAGTGAGGGTTAATTCTGAGCT
TGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATT
CCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGA

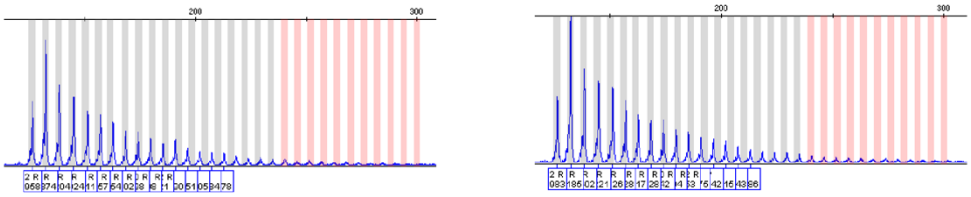
GTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAA
ACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTT
TGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTT
CGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACA
GAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGC
CAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCT
GACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAANNCCGACAGG
ACTATAAAGATACCAGGCGTTTCCCCCTGNAAGCTCCCTCGTGCGCTCTCCTGTT
CCGACCCTGCCGCTTANCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGG
CGCTTTCTCATAGCTCACGCTGTAGGTATCTCNGTTCGGTGTNGNCGTTCGCTCN
AAGCTGGGCTGTGTGNACNAACCCCCCGTTNNANNNNGNACCGNNNNNNNCNTA
NNCNN

(a) Sequencing of G4C2 downstream region

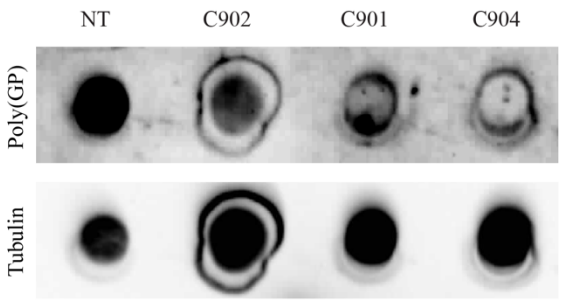
NNNNNNNNNNNNNNNGATTGATTTAGCGGCCGCGAATTCGCCCTTCTCCTGGGA
AAGTGCAGGACCTCCCTCCTGTTTCCGAATACAAAGCCTGGTGGTGTTCAACGCG
GCCAGATAGACCCAATGAGCACACGGACATGTAATCTGTGCACTTCTTTAGACA
ACTGATTACCATCAGTCAAGTGATGCCCAAGTCACAATAGTCACTTCCTTTAAGC
AAGTCTGTGTCATCTCGGAGCTGTGAAGCAACCAGGTCATGTCCCACAGAATGG
GGAGCACACCGACTTGCATTGCTGCCCTCATATGCAAGTCATCACCACTCTCTAG
AAGCTTGGGCTGAAATTGTGCAGGCGTCTCCACACCCCCATCTCATCCCGCATGA
TCTCCTCGCCGGCAGGGACCGTCTCGGGTTCCTAGCGAACCCCGACTTGGTCCGC
AGAAGCCGCGCGCCGCCCACCCTCCGGCCTTCCCCCAGGCGAGGCCTCTCAGTAC
CCGAGGCTCCCTTTTCTCGAGCCCGCAGCGGCAGCGCTCCCAGCGGGTCCCCGGG
AAGGAGACAGCTCGGGTACTGAGGGCGGGAAAGCAAGGAAGAGGCCAGATCCC
CATCCCTTGTCCCTGCGCCGCCGCCGCCGCCGCCGCCGGGAAGCCCGGGGCC
CGGATGCAGGCAATTCCACCAGTCGCTAGAGGCGAAAGCCCGACACCCAGCTTC
GGTCAGAGAAATGAGAGGGAAAGTAAAAATGCGTCGAGCTCTGAGGAGAGCCC
CCGCTTCTACCCGCGCCTCTTCCCGGCAGCCGAACCCCAAACAGCCACCCGCCAG

GATGCCGCCTCCTCACTCACCCACTCGCCACCGCCTGCGCCTCCGCCGCCGCGGG
CGCAGGCACCGCAACCGCAAAGGGCGAATTCGTTTAAACCTGCAGGACTAGTCC
CTTTAGTGAGGNTNATTCTGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGT
GTGAATTGTATCCGCTCACATTCACACACATACGAGCCGGAAGCATAAAGNGTA
AAGCCTGGGNGCCTAATGAGTGAGCTAACTCACATTAANNGCGTTGCGCTCACT
GCCNNCTTTCCANNCGGGAAACCTGNCGNNNNGCTGCNTANNANCNNCCANNNN
NGGGGANAGGNGGNNNNNNATTGGGGNNNNNTTCGCTTCNTCNNNNNTGANTC
NCNNNNNTCGGNNNTNCNNNN

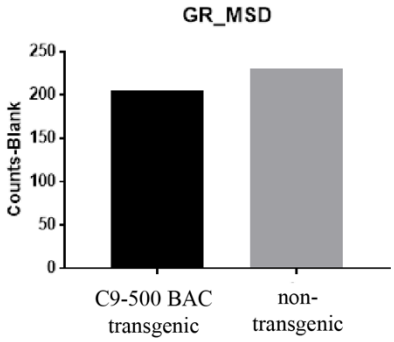
Appendix 4. Repeat-primed PCR traces from additional embryos used.



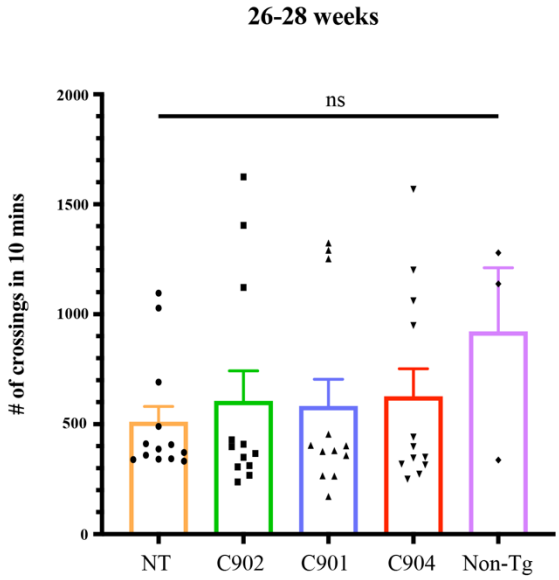
Appendix 5. Additional Dot blot showing decrease in poly-(GP) DPRs upon treatment with C901 or C904.



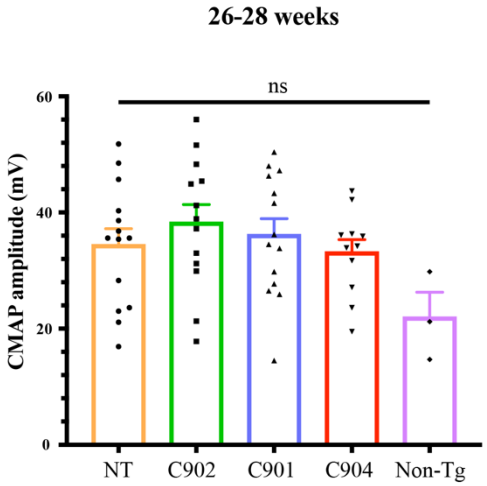
Appendix 6. ELISA assay for poly-(GR) in protein extracts from C9-500 BAC cortical neurons.



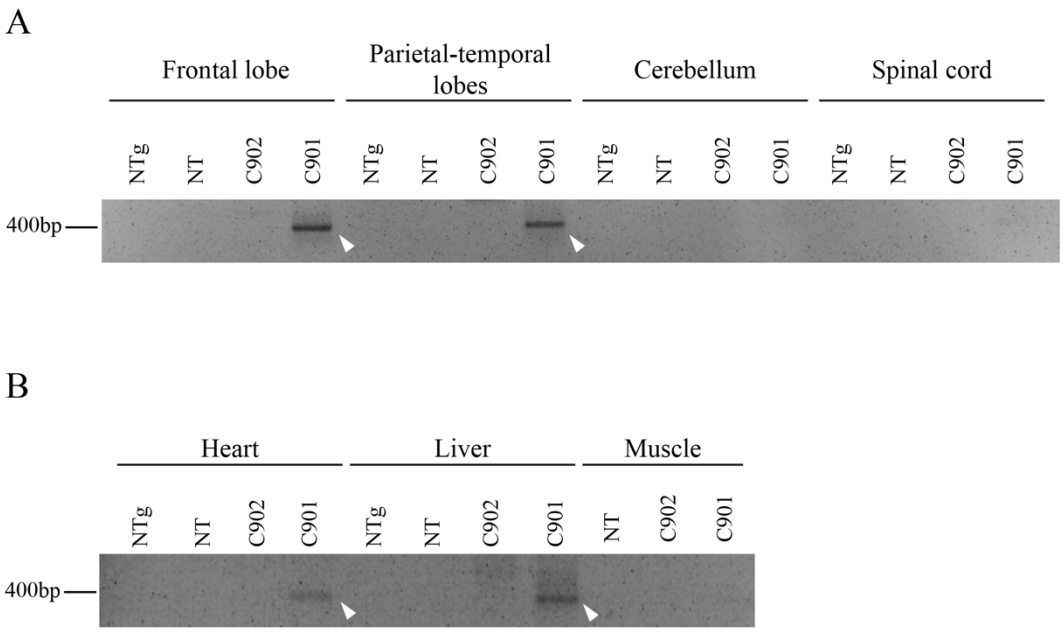
Appendix 7. Open-field test done at a later time-point (26-28 weeks).



Appendix 8. CMAP amplitudes from *in vivo* electrophysiology taken at a later timepoint (26-28 weeks).



Appendix 9. C901-mediated *in vivo* excision of G4C2 repeat expansion from adult C9-500 BAC.



Appendix 10. Excerpt from the Jackson laboratory website (The Jackson laboratory, 2020) regarding their internal investigation into C9-500 BAC phenotype, not peer-reviwed. This is a direct quotation.

“

Please note: The Jackson Laboratory has received inquiries noting that this strain was not demonstrating some of the initial phenotypes reported in the original publication (Liu et al. 2016 Neuron 90:521). In particular that the mice did not develop paralysis. In an independent investigation at The Jackson Laboratory involving over 100 hemizygous females and controls aged more than 15 months, we noted that the mice exhibit some of the key features noted in the publication, including hyperactivity, kyphosis and lethality in a small percentage of mice. However, in our hands, these phenotypes were seen in the same frequency as wildtype (noncarrier) control animals. The nature of the lethalities also appear to correlate with seizures, a trait noted in the FVB/N background. No indication of paralysis were observed in the mice on the study performed at The Jackson Laboratory, and Compound Muscle Action Potential (CMAP) did not vary between hemizygous animals and controls. It should be noted that in the original publication, the penetrance of this phenotype was low. Considerations such as gene by environment interactions can play a part in phenotypic manifestations, and we continue to work closely with the donating investigator and the scientific community to resolve these features in this mouse model. The Jackson Laboratory has confirmed that mice do express the transgene and that dipeptide repeat (DPR) levels of polyGP (as measured by ELISA) are 100X higher than background at two months of age. [2019]

”

Legends for supplementary videos

Supplementary video S1. Video of a C9-500 BAC mouse exhibiting a normal exploratory behaviour in a new environment. This video is shown to highlight the stark difference between littermates with normal and the hyperactive rotating behaviour, which is shown in supplementary video S2.

Supplementary video S2. Video of a C9-500 BAC mouse with a striking hyperactive behaviour characterised by successive runs in circles, always to the same side.

Supplementary video S3. Video showing an example of a transgenic C9-500 BAC mouse with a severe syndrome, exhibiting lethargy, extreme weakness and clasping, but no paralysis.

Supplementary video S4. Video showing a second example of a transgenic C9-500 BAC mouse with a severe syndrome, in this case exhibiting lethargy, extreme weakness and kyphosis but no paralysis.

Supplementary video S5. Video showing an example of a transgenic C9-500 BAC mouse with a mild syndrome, several weeks after suffering from severe symptoms. In the video, the asymptomatic littermates can be seen reacting to stimuli and interacting with the handler and with each other. In contrast, the diseased mouse shows lack of interest in its surroundings, withdrawal from social interactions with its littermates and is unreactive to stimuli, despite not showing obvious signs of distress.

Supplementary video S6. Video showing an example of a non-transgenic littermate with a severe syndrome, exhibiting lethargy, extreme weakness and accelerated/troubled breathing.

Supplementary video S7. Video showing a second example of a non-transgenic littermate with a severe syndrome, exhibiting lethargy and extreme weakness.

Research outputs from PhD

Communications in conferences

Oral platform communications:

João Alves-Cruzeiro, Claudia Bauer, Ian Coldicott, Guillaume Hautbergue, Monika Myszczynska, Laura Ferraioulo, Mimoun Azzouz. Gene editing as a potential therapeutic approach for ALS/FTD-associated with expanded C9ORF72. Presented in Tours, France, at The European Network for the Cure of ALS (ENCALS) conference in May 2019.

João Alves-Cruzeiro. Exploring the role of CRISPR-mediated gene editing as a therapy for C9ORF72-ALS/FTD. Presented in London, UK, at The 2nd European C9ORF72 workshop in January 2020.

Scientific posters:

J Alves-Cruzeiro, E Karyka, C Bauer, I Coldicott, G Hautbergue, C Webster, L Castelli, M Myszczynska, L Ferraioulo, M Azzouz. AAV-mediated gene editing as a potential therapeutic approach for C9ORF72-linked ALS/FTD. Selected for presentation in Barcelona, Spain, at the European Society for Gene and Cell Therapy (ESGCT) conference in October 2019.

João Alves-Cruzeiro, Claudia Bauer, Evangelia Karyka, Ian Coldicott, Guillaume Hautbergue, Christopher Webster, Lydia Castelli, Monika Myszczynska, Laura Ferraioulo, Mimoun Azzouz. Gene editing as a potential therapeutic approach for ALS/FTD-associated with expanded C9ORF72. Presented in Sheffield, UK, at the British Society for Gene and Cell Therapy (BSGCT) conference in June 2019.

Conferences attended and co-chaired:

I have co-chaired the session “Viral and non-viral vector development” and volunteered as a delegate at the British Society for Gene and Cell Therapy (BSGCT) conference in Sheffield, UK, in June 2019.

I have attended the British Society for Gene and Cell Therapy (BSGCT) conference in London, UK, in November 2018.

I have attended and co-chaired the session "Introduction to Sheffield Translation Neuroscience" in the Sheffield Neuroscience Conference, July 2017.

I have attended the British Society for Gene and Cell Therapy (BSGCT) conference in Cardiff, UK, in April 2017.

Scientific publications

Jayanth Chandran, Paul Sharp, Evangelia Karyka, **João Miguel da Conceição Aves-Cruzeiro**, Ian Coldicott, Lydia Castelli, Guillaume Hautbergue, Mark Collins, and Mimoun Azzouz. 2017. Site Specific Modification of Adeno-Associated Virus Enables Both Fluorescent Imaging of Viral Particles and Identification of the Capsid Interactome. Scientific Reports.

Status: published

Alves-Cruzeiro J, Webster CP, Azzouz M. 2019. The hybrid AAVP tool gets an upgrade. PNAS. Commentary article.

Status: published

João Alves-Cruzeiro, Evangelia Karyka, Christopher P Webster, Ian Coldicott, Claudia Bauer, Sneha Simon, Michela Pulix, Adrian Higginbottom, Guillaume Hautbergue, Laura Ferraiuolo, Mimoun Azzouz. Exploring a highly targeted gene editing approach as a potential therapy for C9ORF72-ALS/FTD.

Status: in preparation

<u>Articles accepted for publication during PhD programme but originated from previous work:</u>

Marcelo A, Brito F, Carmo-Silva S, Matos CA, **Alves-Cruzeiro J**, Vasconcelos-Ferreira A, Koppenol R, Mendonça L, de Almeida LP, Nóbrega C. 2019. Cordycepin activates autophagy through AMPK phosphorylation to reduce abnormalities in Machado-Joseph disease models. Human Molecular Genetics.

Status: published

Alves-Cruzeiro, J. M. D. C., Mendonça, L., Pereira de Almeida, L., and Nóbrega, C. 2016. Motor Dysfunctions and Neuropathology in Mouse Models of Spinocerebellar Ataxia Type 2: A Comprehensive Review. Frontiers in Neuroscience.

Status: published