Ricardo António Afonso Reis

Silencing *ATXN2* through CRISPR-mediated insertion of a premature STOP codon

Faculty of Medicine and Biomedical Sciences

2020

Ricardo António Afonso Reis

Silencing ATXN2 through CRISPR-mediated insertion of a premature STOP codon

Master's degree in Biomedical Sciences – Disease mechanisms

This work was done under the supervision of:

Carlos Matos, PhD

Clévio Nóbrega, PhD

University of Algarve

Faculty of Medicine and Biomedical Sciences

2020

Authorship Statement

I hereby declare to be the author of this work, which is original and unpublished. Authors and papers consulted are duly cited in the text and are listed in the included references.

(Ricardo António Afonso Reis)

Index of Contents

Acknowledgments/ Agradecimentos

Um simples obrigado não descreve o quanto me sinto agradecido ao Doutor Clévio Nóbrega, que desde sempre apesar da sua posição tanto a nível institucional com a nível de científico, me tratou como o cientista que quero um dia poder vir a ser. Obrigado por me receberes de braços abertos nesta grande família que é o grupo MNGT. Ter um PI que transmite uma energia tão radiante e que é de tal maneira acessível que nunca existe pergunta uma má pergunta, é de facto um privilégio. Quero ainda agradecer-te por me teres dado a conhecer esta área pela qual me tenho vindo a interessar cada vez mais, sem ti estaria possivelmente a trabalhar noutra área científica e não seria tão feliz quanto sou agora.

Penso que deve ser difícil para todas as pessoas escrever a secção em que se agradece ao orientador principal, mas para mim, conseguir encontrar por onde começar a agradecer ao Doutor Carlos Matos está a revelar-se extremamente difícil. Isto porque o Carlos não foi apenas um orientador académico, mas também um orientador da vida. Vou carregar pela minha vida fora os princípios que me foram transmitidos por ti. Por isso um obrigado. Agradeço também a paciência para me explicares ao pormenor todos os aspetos necessários para compreender a fundo qualquer questão que pudesse surgir. Além de todo o conhecimento que me transmitiste, tenho a certeza de que no futuro muito mais conhecimento está para vir, não fosses tu um repositório de conhecimento sem fundo nas mais diversas áreas. No lugar de um orientador encontrei um colega e um amigo que me indicava o caminho e me dava a ferramentas para alcançar o objetivo com as minhas próprias mãos. Esta necessidade de pensar por mim próprio sobre o teu olhar atento fez-me crescer imenso tanto como pessoa, como profissional. Isso não tem preço. Quando comecei a minha estadia no laboratório não sabia o que esperar da nossa relação, agora que esta dissertação está entregue posso dizer que saí com mais um amigo. Por esta dualidade profissional e amistosa estou-te muito grato. Espero que no futuro mais pessoas fiquem na tua alçada, e acabem esta etapa tão gratos quanto eu.

Quero deixar um grande obrigado aos atuais estudantes de doutoramento do laboratório, André Conceição, Adriana Marcelo e Rebekah Koppenol, que se disponibilizaram sempre que possível para me ensinar e ajudar no possível e impossível. Contudo não posso deixar que frisar o contributo especial que a Rebekah teve para a realização do meu projeto, foi realmente incansável em todos os dias e a toda a hora. Não posso deixar e fora os meus restantes colegas de laboratório MNGT, um obrigado a todos por tornarem este grupo tão dinâmico e tão eletrizante. Agradeço ainda aos nossos colegas do lado, especialmente ao Leonardo que sempre

que necessitei me estendeu a mão. A todos estes colegas de batalha diária um imenso obrigado pela amizade e por tornarem o centro mais animado.

Um obrigado enorme à Inês Afonso que desde a nossa entrada na universidade me atura e ajuda de tal maneira que nunca serei capaz de lhe pagar.

Ao meu advogado privado e eterno amigo Tomás Nunes um imenso obrigado por todo o apoio que me tem dado pela vida fora.

Sem dúvida um dos meus pilares emocionais durante a realização deste projeto, foi a Cristiana Madeira. Obrigado por me aturares todos os dias, quando as coisas correram bem e estava feliz e quando as coisas correram mal e estava desgastado. Para ti um eterno obrigado por acreditares em mim e me motivares para seguir em frente.

Um enorme obrigado à minha tia Isabel e ao meu tio Raul que são como uns segundos pais para mim. Obrigado por todo o apoio e carinho.

Ao meu padrasto Amandio Farias um agradecimento sem fim por tudo o que fez e continua a fazer. És como um pai para mim.

Em último lugar quero agradecer à minha mãe Maria Irene que nunca pensou duas vezes em se sacrificar para que eu conseguisse alcançar mais uma etapa. Sem o esforço infinito dela nunca seria possível chegar onde estou de momento. Devo-lhe tudo o que sou e o que virei a ser. Para ela um eterno obrigado.

Abstract

Spinocerebellar ataxia type 2 (SCA2) is a hereditary neurodegenerative disorder caused by expansion of CAG trinucleotide repeats present in the codifying region of the *ATXN2* gene. The mutant *ATXN2* product - a ataxin-2 protein with an expanded polyglutamine tract - displays a cytotoxic gain-of-function, leading to progressive neurodegeneration. No definitive therapy for this disease has yet been developed, but *ATXN2* knock-down represents a promising approach. The versatility of the CRISPR-Cas system enables gene silencing at a genomic level, through an array of unprecedently diverse mechanisms. Cas nucleases guided by a single guide RNA (sgRNA) molecule produce modifications in precise regions of the DNA, and the system can be modified to direct transcriptional inhibitors such as the Krüppel-associated box domain (KRAB) to particular genetic loci. The aim of this study was to develop two *ATXN2* silencing strategies, utilizing catalytically active Cas9 to knock-down the *ATXN2* gene or, alternatively, a dCas9-KRAB fusion protein to pre-transcriptionally repress *ATXN2* expression. One sgRNA targeting the codifying region of ATXN2 was cloned into a plasmid encoding Cas9, and another sgRNA targeting the *ATXN2* promoter was cloned into a plasmid encoding dCas9-KRAB. Plasmids were transfected into human embryonic kidney (HEK) 293T cells and endogenous *ATXN2* mRNA and protein levels were analyzed by qPCR and Western blot. Cas9/sgRNA did not alter *ATXN2* mRNA levels in comparison to controls, contrasting with ataxin-2 protein levels that revealed a decrease tendency. dCas9-KRAB/sgRNA did not alter *ATXN2* mRNA levels, but displayed a tendency for decreasing ataxin-2 protein levels, relative to non-transfected cells. Our study suggests that the CRISPR-Cas9 system may be adapted for *ATXN2* gene silencing applications, but further improvements to our strategies will be necessary to produce significant alterations.

Keywords: Spinocerebellar ataxia type 2 (SCA2), polyglutamine diseases, neurodegeneration, gene editing, CRISPR/Cas system, gene silencing

Resumo

De entre as diferentes formas hereditárias de ataxia, a ataxia espinocerebelosa tipo 2 representa a segunda mais comum a nível mundial. Os sintomas progressivos desta doença neurodegenerativa incluem perda de coordenação motora, movimentos sacádicos oculares e vários sinais resultantes de degeneração cerebelosa, como disartria e disfagia. A ataxia espinocerebelosa tipo 2 surge como consequência de uma expansão anormal de repetições de trinucleótidos CAG presentes no exão 1 do gene *ATXN2*, em número igual ou superior a 33. O gene *ATXN2* codifica a proteína ataxina-2 e, na sua forma mutada, origina uma forma alterada desta proteína, com uma sequência de glutaminas anormalmente expandida. Esta expansão confere à proteína propriedades citotóxicas, responsáveis pela disfunção e morte de populações particulares de neurónios, nomeadamente as células de Purkinje. Com efeito, a ataxia espinocerebelosa tipo 2 insere-se no grupo das doenças de poliglutaminas, um conjunto de nove doenças neurodegenerativas monogénicas que têm em comum o facto de serem causadas por expansões de repetições de glutaminas nas proteínas a que estão associadas. Recentemente, estudos têm sugerido que tratos de CAG expandidos presentes em transcritos de mRNA podem também apresentar toxicidade e contribuir para os mecanismos fisiopatológicos.

Até aos dias de hoje, não foi ainda desenvolvida nenhuma terapia eficaz para a ataxia espincerebelosa tipo 2, que seja capaz de tratar a doença ou de atrasar o seu curso. Como tal, o tratamento da ataxia espinocerebelosa tipo 2 tem sido limitado ao controlo dos sintomas. Algumas estratégias terapêuticas estão atualmente a ser desenvolvidas e diversos alvos terapêuticos continuam por explorar. Algumas das estratégias mais promissoras baseiam-se no silenciamento do gene *ATXN2*, mas até hoje ainda não foram testados métodos de silenciamento que atuem a nível genómico, em contraste com os métodos clássicos que atuam sobre as moléculas de ARNm.

O sistema de edição génica CRISPR/Cas9 tem expandido as possibilidades de manipulação que pode ser exercida sobre o genoma, e sobre a expressão genética. Este sistema tem como base uma endonuclease, Cas9, e uma molécula de ARN guia (*single guide RNA*; sgRNA) que, quando combinadas, formam um complexo ribonucleoproteíco. O ARN guia direciona o complexo para uma região alvo do ADN, ligando-se a essa sequência por complementaridade de bases. A proteína Cas9 tem a capacidade de clivar ambas as cadeias da hélice dupla de ADN, introduzindo assim um corte na cadeia dupla. Este sistema pode ser explorado para produzir mutações em regiões particulares do genoma, decorrentes da reparação

da região clivada. Alternativamente, a proteína Cas9 pode ser modificada e usada para direcionar inibidores da transcrição, como o domínio *Krüppel-associated box* (KRAB), para loci genéticos específicos.

Este projeto teve como objetivo desenhar e desenvolver uma estratégia com potencial terapêutico para a ataxia espinocerebelosa tipo 2, sustentada no silenciamento do gene *ATXN2* e recorrendo a ferramentas moleculares baseadas no sistema CRISPR/Cas9. Para alcançar este objetivo, foram testadas duas abordagens distintas. Primeiramente, procurou-se utilizar o sistema CRISPR/Cas9 convencional de modo a introduzir um codão de terminação prematuro no gene *ATXN2* e assim incapacitar permanentemente a tradução de ataxina-2, num fase inicial da síntese proteica. Numa estratégia alternativa, recorreu-se a uma variante do sistema CRISPR/Cas, designado de *CRISPR interference* (CRISPRi), para reprimir a expressão do gene *ATXN2* a um nível pré-transcricional. Esta última abordagem tomaria partido de um complexo proteico constituído por uma proteína Cas9 sem atividade catalítica (dCas9), fundida com um domínio KRAB.

O bloqueio da expressão genética mediada por CRISPR/Cas9, na sequência de um corte da cadeia dupla de DNA, depende de um sistema endógeno de reparação do ADN - a união terminal não homóloga (*non-homologous end-joining*; NHEJ). Este processo introduz pequenas inserções ou deleções (*indels*) na cadeia de ADN, que levam consequentemente à formação de um codão de terminação prematuro a jusante do local reparado. De modo a introduzir um corte na cadeia dupla de DNA a montante do trato de CAGs do gene *ATXN2*, desenhou-se um sgRNA complementar a uma sequência presente entre o codão de iniciação da tradução e o início das repetições de CAG. A sequência de sgRNA foi clonada num plasmídeo de expressão que codificava Cas9 cataliticamente ativa, e este plasmídeo (Cas9/Indel) foi então transfetado em células HEK 293T, com o objectivo de se determinar o seu efeito na expressão do gene *ATXN2*, a nível de ARNm e da proteína ataxina-2, através de qPCR e Western blot, respetivamente.

Os resultados de qPCR indicaram que os níveis de expressão de ARNm do gene *ATXN2* em células HEK 293T transfetadas com Cas9/Indel não apresentavam uma diferença estatisticamente significativa comparativamente às condições controlo. Quanto aos resultados de Western blot, as células transfetadas com a ferramenta molecular em teste apresentaram uma tendência para uma redução dos níveis proteicos de ataxina-2. Contudo, tal como nos resultados de qPCR, esta diferença não se demonstrou estatisticamente significativa.

Com o objetivo de silenciar a expressão de *ATXN2* através de CRISPRi, desenharam-se várias sgRNAs capazes de direcionar a proteína de fusão dCas9-KRAB para o promotor do gene *ATXN2*. Esperava-se que a ligação da proteína dCas9-KRAB ao promotor do gene *ATXN2* silenciaria a expressão do gene através do recrutamento, mediado pelo KRAB, de maquinaria modificadora de cromatina. Umas das sequências de sgRNAs foi clonada num plasmídeo codificando a fusão dCas9-KRAB, e células HEK 293T foram transfetadas com o plasmídeo assim gerado (dCas-KRAB/Prom2). O efeito desta estratégia na expressão do gene *ATXN2* foi avaliado novamente a nível do mARN e da proteína ataxina-2, através de qPCR e Western blot.

Os resultados de qPCR relativos aos níveis de ARNm do gene *ATXN2* expressos por células transfetadas com dCas-KRAB/Prom2 não revelaram diferenças estatisticamente significativas, em comparação com os restantes grupos experimentais. A análise dos resultados de Western blot revelou uma tendência para a redução dos níveis proteicos de ataxin-2, comparando com as células não transfetadas, embora, uma vez mais, sem significância estatística.

Este estudo demonstra a variedade de estratégias de silenciamento genético que se podem elaborar com base no sistema CRISPR/Cas9. Contudo, serão necessários estudos adicionais e a optimização de alguns dos procedimentos utlizados antes que se possa validar e estabelecer o potencial das duas estratégias desenvolvidas neste trabalho como estratégias putativas para o silenciamento do gene *ATXN2* e o tratamento da ataxia espinocerebelosa tipo 2.

Palavras-chave: Ataxia espinocerebelosa tipo 2, neurodegeneração, edição génica, sistema CRISPR/Cas9, silenciamento génico

List of Figures

List of tables

List of abbreviations

4-HT - 4-hydroxytamoxifen

A2BP1 - ataxin-2 binding protein 1

AAV - Adeno-associated viral

AD - Activation domain

AID - Activation-induced cytidine deaminase

ASOs - Antisense oligonucleotides

BDNF - Brain-derived neurotropic factor

CAG – Cytosine-Adenine-Guanine

Cas - CRISPR-associated

Cdk 5 - Cyclin-dependent kinase 5

CHZ - Chlorzoxazone

CRISPR - Clustered regularly interspaced short palindromic repeat

CRISPRi – CRISPR interference

crRNA - CRISPR RNA

CUG - Cytosine-Uracil-Guanine

CXCR4 - C-X-C chemokine receptor type 4

Cyt C - cytochrome C

DDX6 - DEAD/H-box RNA helicase 6

DHFR - dihydrofolate reductase

Dnmt - DNA methyltransferases

DRPLA - Dentatorubral-pallidoluysian atrophy

DSB - double-strand break

dsRNA - double-stranded RNA

ER - Endoplasmic reticulum

FKBP - FK506 binding protein

FUS - Fused in sarcoma

GDNF - Glial cell-derived neurotrophic factor

GFP - Green fluorescent protein

HD - Huntington's disease

HDR - Homology-directed repair

HEK - Human embryonic kidney

HR - Homologous recombination

Hsp40 - Heat shock proteins 40

Hsp70 - Heat shock proteins 70

Htt - huntingtin

Indel - insertion and/or deletion

IP3R - Inositol-3- trisphosphate receptor

iPSC - induced pluripotent stem cells

KRAB - Krüppel-associated box

Lsm - Like Sm domain

LsmAD - Lsm-associated domain

miRNA - microRNA

MJD - Machado-Joseph disease

MSCs - Multipotent stromal fibroblast-like cells

NAT - Natural antisense transcript

NGG - Nucleotide-Guanine-Guanine

NHEJ - nonhomologous end-joining

NLS - Nuclear localization signal

PABP - Poly(A)-binding protein

PAM - Protospacer adjacent motif

PAM2 - Polyadenylate-binding protein 1 interacting motif-2

p-bodies - Processing bodies

PC - Purkinje cell

PCK - photocaged lysine

polyQ – Polyglutamine

PINK1- PTEN Induced Kinase 1

RBD - RNA binding domains

RBP - RNA binding protein

RISC - RNA-induced silencing complex

RNAi - RNA interference

RVD - Repeat-variable di-residue

SBM2 - Splicing binding motif 2

SBMA - Spinal and bulbar muscular atrophy

SCA1 - Spinocerebellar Ataxia Type 1

SCA17 - Spinocerebellar Ataxia Type 17

SCA2 - Spinocerebellar Ataxia Type 2

SCA3 - Spinocerebellar Ataxia Type 3

SG - Stress granules

sgRNA - single-guide RNA

SH3 - Sarcoma homology 3

shRNA - short hairpin RNA

siRNA - small interfering RNA

SK channel - Small-conductance calcium-activated potassium channels

SNPs - single-nucleotide polymorphisms

SpCas9 - *S. pyogenes* Cas9

STAU1 - Staufen1

TadA - tRNA adenine deaminase

TALE - Transcription activator-like effector

TALEN - Transcription activator-like effector nuclease

TDBS - TALE DNA-binding scaffold

TDP-43 - TAR DNA binding protein 43

TMP – Trimethoprim

tracrRNA - trans-activating crRNA

ZFN - Zinc finger nuclease

ZFP- Zinc finger proteins

CHAPTER 1 – INTRODUCTION

1.1. Spinocerebellar Ataxia Type 2 (SCA2)

1.1.1. Brief overview of polyglutamine diseases

Polyglutamine (polyQ) diseases are a group of nine rare neurodegenerative disorders that includes Huntington's disease (HD), dentatorubral-pallidoluysian atrophy (DRPLA), spinal and bulbar muscular atrophy (SBMA), and six forms of autosomal dominant spinocerebellar ataxia (SCAs): SCA1, 2, 3, 6, 7 and 17. This disorders arise from abnormal expansion of *cytosine-adenine-guanine* (CAG) repeats in the coding region of particular genome loci, which codify a tract of repeated glutamine residues (Nóbrega & Pereira de Almeida 2018). Despite these genes being otherwise unrelated and the fact that the resulting proteins do not share significant homology apart from the CAG/polyQ tract, polyQ-expanded proteins induce cellular toxicity and tend to aggregate, resulting in the formation of insoluble protein aggregates in the central nervous system, a key feature of polyQ disorders. Different polyQ diseases are associated with distinct clinical and neuropathological features, but SCAs share the fact that they arise from cerebellar deficits accompanied by ataxia, a decrease of motor coordination (Buijsen et al 2019, Nóbrega & Pereira de Almeida 2018).

1.1.2. Spinocerebellar ataxia type 2 epidemiology

Spinocerebellar ataxia type 2 (SCA2) is the second most common form of SCA worldwide, affecting 0,1 to 5,8 per 100.000 inhabitants, and also one of the most severe types (Lorenzetti et al 1997). Global prevalence evaluation studies showed large numbers of SCA2 families in Mexico, India, South Africa, Venezuela, and Italy (Velázquez-Pérez et al 2017).

Cuba presents the highest worldwide prevalence rate of SCA2, with 6,57 cases per 100.000 inhabitants and a frequency of mutation carriers of 28,51 per 100.000 inhabitants, due to a prominent founder effect. Cuba also represents que nation with the highest frequency of large *ATXN2* normal alleles, which leads to a continuous rise in SCA2 prevalence rate, as a result of these normal alleles being unstable and thus prone to expansion (Auburger et al 1990, Laffita-Mesa et al 2014). The highest frequencies within Cuba are observed in the northern region of the island, in the province of Holguin, with 40,18 cases and 182,75 carriers per 100.000 inhabitants. Remarkably, municipalities within Holguin province present extremely

high incidence rates, with around 18 annual new cases per 100.000 inhabitants (Velazquez Perez et al 2009).

1.1.3. SCA2 characterization

1.1.3.1. Clinical features

Similarly to other SCAs, SCA2 is characterized by a progressive cerebellar syndrome which usually culminates in the hallmark characteristic and most noticeable symptom of SCA2: gait ataxia. Ataxia onset is frequently accompanied by muscle cramping and followed by multiple other symptoms resulting from cerebellar degeneration. These symptoms include appendicular ataxia with instability of stance, dysarthria, and oculomotor deficits such as nystagmus and ocular dysmetria (Buijsen et al 2019, Scoles & Pulst 2018). Although SCA2 symptomatology is almost entirely due to cerebellar degeneration, slow or absent horizontal saccadic eye movements, which is the most noticeable oculomotor disturbance in SCA2 and also a pathognomonic sign of the disease, arises from neuron degeneration in the oculomotor brainstem. Other non-cerebellar manifestations of SCA2 such as dystonia, myoclonus, neuropathy, muscle spasticity and frontal-executive dysfunction are also frequent. Furthermore, SCA2 patients with variant phenotypes may present idiopathic forms of parkinsonism or amyotrophic lateral sclerosis (ALS) (Ashizawa et al 2013, Geschwind et al 1997, Velázquez-Pérez et al 2017). Notably, toxic damage to neuronal structures starts up to 15 years before the ataxic onset, manifesting through subtle motor and unspecific non-motor features (Velázquez-Pérez et al 2017).

Despite SCA2 being considered a late-onset disease, with initial symptoms appearing around the fourth decade of life, infantile or pediatric cases of SCA2 have been described in some populations with very large expansions of CAG repeats. The described clinical features of infantile SCA2 include severe cerebellar impairment, in conjunction with a large set of rare symptoms and signs such as myoclonus-epilepsy, retinitis pigmentosa, tetraparesis, dysphagia, cognitive impairment, facial dysmorphism, developmental delay, oculomotor apraxia, deterioration of expressive language, comprehension and memory, graphomotor skills and dysarthria, in addition to progressive extrapyramidal manifestations, dysphagia, trophic changes and slow eye movements (Babovic-Vuksanovic et al 1998, Di Fabio et al 2012, Gigante et al 2020, Magaña et al 2013, Ramocki et al 2008).

1.1.3.2. Neuroanatomical features

Macroscopic observations of post-mortem SCA2 brain samples revealed significant atrophy of the cerebellum, brainstem, cranial nerves, and frontal lobe, which results in an overall reduction of brain size. Furthermore, with great relevance to this disorder, white matter in the cerebral and cerebellar regions showed detectable reduction, in addition to midbrain substantia nigra discoloration. Histopathological studies showed an untimely and pronounced loss of neurons from the cerebellar Purkinje cell layer, accompanied by reduced dendritic arborizations and torpedo-like deformations of axons. Brainstem samples displayed degeneration of pre-cerebellar brainstem nuclei and considerable loss of inferior olive neurons. Widespread neuronal loss was observed in the cerebral cortex, thalamus, basal forebrain, spinal cord and in the substantia nigra from the mesencephalon (Estrada et al 1999, Gierga et al 2005, Rub et al 2007, Seidel et al 2012).

1.1.4. Genetics of SCA2

SCA2 was first described in 1971, with the discovery of nine affected Indian families (Wadia & Swami 1971)**,** but it wasn´t until 1996 that the causative mutation in the *ATXN2* gene (NG_011572.3) was identified as an unstable expansion of the CAG repeat region in the exon 1 of *ATXN2*. This mutation culminates in the translation of a protein - Ataxin-2 - with a polyglutamine expansion (Pulst et al 1996, Sanpei et al 1996)**.** The *ATXN2* gene is located at the 12q24-12 chromosomal region, is constituted by 25 exons (**Figure 1A**) and spans a total of 147.463 bp. The resulting transcript, with 4699 bp, has two in-frame start codons located at the 5´-end of the sequence, with the second one being four codons upstream of the CAG tract. Intriguingly the identity of the one that initiates translation is still a matter of debate (Scoles & Pulst 2018). Ataxin-2 (NP_001359503.1) is a 144 kDa protein with a glutamine-rich region of variable length in the amino-terminal domain; the length of polyglutamine tract increases with the size of its coding CAG-repeat region.

Normal alleles of the *ATXN2* gene carry between 13 and 31 CAG trinucleotide repeats, with a 22 CAG repeats allele being the most common allele described worldwide. Genetic stability of normal alleles, due to low mutation rates, appears to be conferred by the presence of one or two alternative glutamine coding triplets, namely CAA, in between CAG repeated units [$(CAG)_8$-(CAA)-$(CAG)_4$-(CAA)-$(CAG)_8$] (Magaña et al 2013, Velázquez-Pérez et al

2017). Interestingly, alleles with repeats interrupted by CAA or with 28-33 long CAG expansions appear to predispose carriers to develop a Parkinson's disease-like phenotype. Furthermore, alleles carrying expansions with 28-33 repeats are also considered to increase risk for ALS (Ross et al 2011, Yu et al 2011).

ATXN2 alleles with 33 to 500 CAG repeats are the ones associated with the development of SCA2. These alleles are highly unstable during meiosis, since they do not include CAA interruptions and their absence increases the propensity for further expansion of the CAG tract. SCA2 is characterized by the "anticipation" phenomenon common to other polyQ disease, whereby an earlier disease onset and more severe manifestations strongly correlated with an increase of CAG repeat length observed throughout successive generations. The aforementioned infantile or pediatric cases of SCA2 are associated with very large CAG repeat numbers, of 130 to ≥ 200 (Imbert et al 1996, Scoles & Pulst 2018).

Noteworthily, contrary to other polyQ disorder-associated genes, *ATXN2* alleles do not present a range of intermediate CAG repeat with clinical significance. Although not associated with SCA2 disease development, these repeats are nonetheless prone to further expansion into the pathological range (Magaña et al 2013, Matsumura & Futamura 2001).

1.1.5. Ataxin-2 functions

The ataxin-2 protein contains several different domains: the polyQ tract, a Like Sm domain (Lsm), a Lsm-associated domain (LsmAD), a Sarcoma homology 3 domain (SH3) with a splicing binding motif (SBM2), and a Polyadenylate-binding protein 1 interacting motif-2 (PAM2), located next to the carboxyl-terminal (**Figure 1B**) (Albrecht et al 2004, Satterfield & Pallanck 2006, Scoles & Pulst 2018). Ataxin-2 is a ubiquitously expressed cytoplasmic protein, found in the rough endoplasmic reticulum (ER), Golgi complex, RNA-protein granules and, in a pathological context, inclusion bodies. This protein is involved in a vast multitude of cellular processes, such as RNA metabolism (at the levels of alternative splicing and protein translation), nutritional and energetic sensing, SG and processing bodies formation, calcium homeostasis maintenance and endocytosis. To better understand the role of ataxin-2 in these processes, it is important to consider the proteins with which it interacts and their respective functions (Egorova & Bezprozvanny 2019, Scoles & Pulst 2018).

The ataxin-2 binding protein 1 (A2BP1/RBFOX1) is implicated in the regulation of RNA alternative splicing. ataxin-2 was found to interact with A2BP1, in addition to interacting directly with RNA through the Lsm and LsmA domains (**Figure 2A**). Considering that A2BP1 is not ubiquitously expressed, it's binding to ataxin-2 was suggested to have a tissue-specific role in regulating alternative splicing of specific RNAs (Satterfield & Pallanck 2006, Shibata et al 2000).

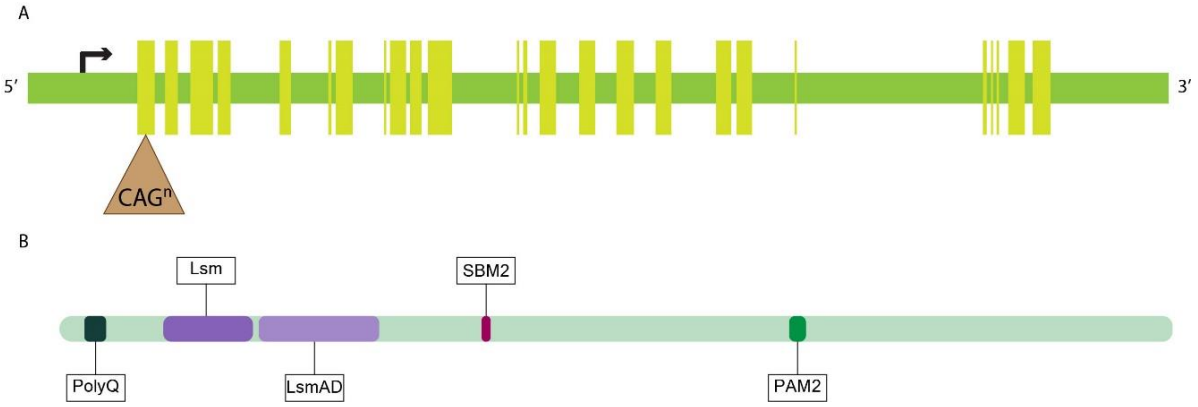

Figure 1- *ATXN2* gene structure and Atxn2 protein scheme. A) The *ATXN2* gene comprises the 25 exons (light green boxes represent), encoding the Atxn2 protein. The arrow denotes the transcription start site. The CAG region of variable length, that when expanded beyond 31 repeats leads SCA2, is localized in exon 1 (represented by brown triangle). **B)** The ataxin-2 protein is constituted by 1.312 amino acids and has 5 domains identified so far. These domains include: the polyQ tract (aa 6-27), Like Sm domain (Lsm; aa 94-185), Lsm-associated domain (LsmAD; aa 193-315), SRC homology 3 domain with a splicing biding motif 2 (SBM2; aa 427-436) and a poly(A)-binding protein (PABP) interacting domain 2 (PAM2; aa 748-765).

Stress granules are nuclear and most commonly cytoplasmic foci formed by the condensation of untranslated mRNAs and RNA-binding proteins (RBPs), that are assembled as a response to cell stress and help cells withstand deleterious conditions (Khong et al 2017). Poly(A)-binding protein (PABP), a component of mammalian SGs, interacts with the poly(A) end of mRNAs during the initiation of translation. It has been demonstrated that ataxin-2 and PABP colocalize and physically interact. This interaction has been suggested to take place via the PAM2 domain of ataxin-2 (**Figure 2A/B/C**). Moreover, a reduction in ataxin-2 levels leads to increased levels of PABP and abnormal SG formation. These results reinforce the role of ataxin-2 in translation regulation and demonstrate its presence in SGs (Nonhoff et al 2007, Ralser et al 2005a). Interestingly, ataxin-2 has been proposed to act in a context of energy deficit as a nutritional and energetic sensor, acting with PABP to control translation of proteins

involved in the mTOR signaling pathway, a central regulator of eukaryotic cell growth and metabolism (Becker et al 2017).

Further confirmation that ataxin-2 is involved in SGs formation was achieved by demonstrating that ataxin-2 interacts through the Lsm and LsmAD domains with the DEAD/H-box RNA helicase (DDX6) (**Figure 2C**), a protein localized in SGs. Considering that DDX6 is also found in processing bodies (p-bodies), cytoplasmic granules containing repressed mRNAs and proteins related to mRNA decay, ataxin-2 is suggested to also play a role in p-body formation (Nonhoff et al 2007).

Staufen1 (STAU1) is a double-stranded RNA-binding protein that acts as a key regulator of the formation of SGs and cytoplasmic inclusions in neuroglia, by targeting specific mRNAs for degradation. This protein was also identified to interact directly with ataxin-2, suggesting that ataxin-2, together with STAU1, plays a functional role in SGs by mediating the processing of SG mRNAs (**Figure 2C**) (Paul et al 2018)

The RNA-binding domains (RBDs) of TAR DNA-binding protein (TDP-43) and fused in sarcoma protein (FUS) are mutated in some forms of ALS. Ataxin-2 was demonstrated to interact with both TDP-43 and FUS in an RNA-dependent fashion (Elden et al 2010, Farg et al 2013). Furthermore, mutant TDP-43 and FUS have been observed in SGs carrying ataxin-2 (**Figure 2C**). Taken together with evidence that ataxin-2 with intermediate-length polyQ expansions increases the risk for developing ALS, these observations suggest that the increased ALS risk associated with ataxin-2 is due to abnormal SG function (Elden et al 2010, Nihei et al 2012).

Rgs8 is a regulator of G-protein signaling thought to regulate calcium homeostasis in Purkinje cells via mGluR1 inhibition. Its mRNA was determined to interact with ataxin-2 (**Figure 2D**) in addition to Rsg8 protein levels being shown to be reduced in the presence of mutant ataxin-2 (Dansithong et al 2015). Moreover, inositol trisphosphate receptor (IP3R), responsible for mediating calcium release and intake, was shown to turn abnormally activate upon interacting with mutant ataxin-2 but not with non-expanded ataxin-2 (**Figure 2D**). Taken together, these interactions support the role of ataxin-2 in calcium homeostasis maintenance, by mediating calcium intake and release (Liu et al 2009).

Although mutated parkin is responsible for the development of Parkinson's disease, Parkinson's disease-like phenotypes are observed in some cases of SCA2. In studies trying to address this connection, both wild-type and mutated ataxin-2 were shown to be ubiquitinated

by parkin, a E3 ubiquitin-ligase, involved in the reorganization of the actin cytoskeleton during endocytosis. This results suggest a direct interaction between parkin and ataxin-2 (**Figure 2E**). In the case of parkin being mutated, abnormally expanded ataxin-2 ubiquitination was demonstrated to be less efficient, which leads to the decreased turnover of the polyQ protein. Although the relation between SCA2 and Parkinson´s disease is not yet fully understood, the role of parkin in regulating the levels of available mutant ataxin-2 may give an insight into the connection between SCA2 and the development of Parkinson's disease-like phenotype (Huynh et al 2007).

Endophilins are proteins that function together in endocytosis of cell surface receptors. ataxin-2 was demonstrated to bind directly to both endophilin A1 and A3 through the SBM2 domain (Ralser et al 2005b). Furthermore, endocytosis complexes containing endophilin A3, CIN85 and Cbl were shown to also contain ataxin-2 (**Figure 2E**) (Nonis et al 2008). These finding suggest that ataxin-2 is involved in endocytosis, although a specific role has not been established (Soubeyran et al 2002).

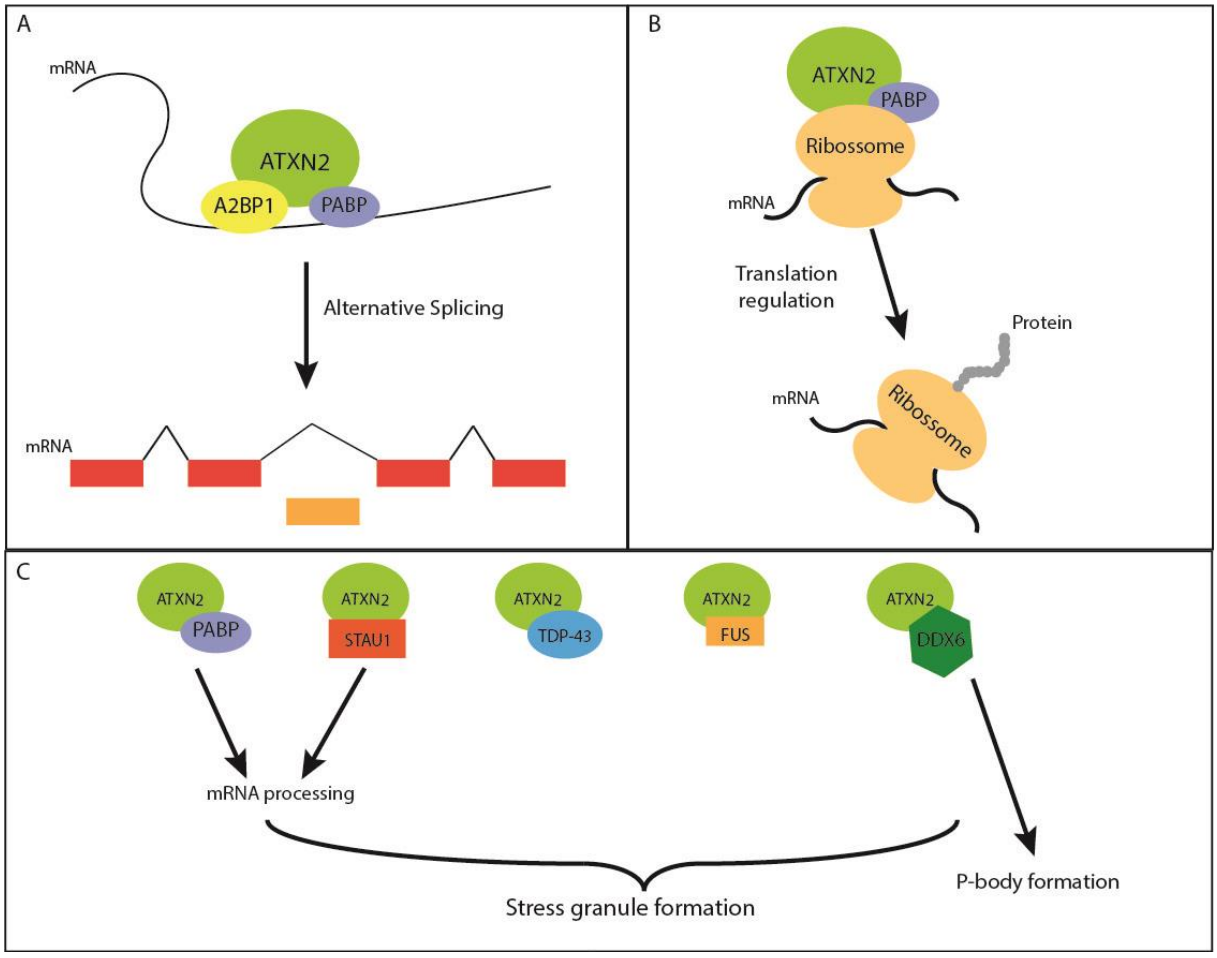

Figure 2- Different cellular processes in which ataxin-2 is implicated (Figure continues below)

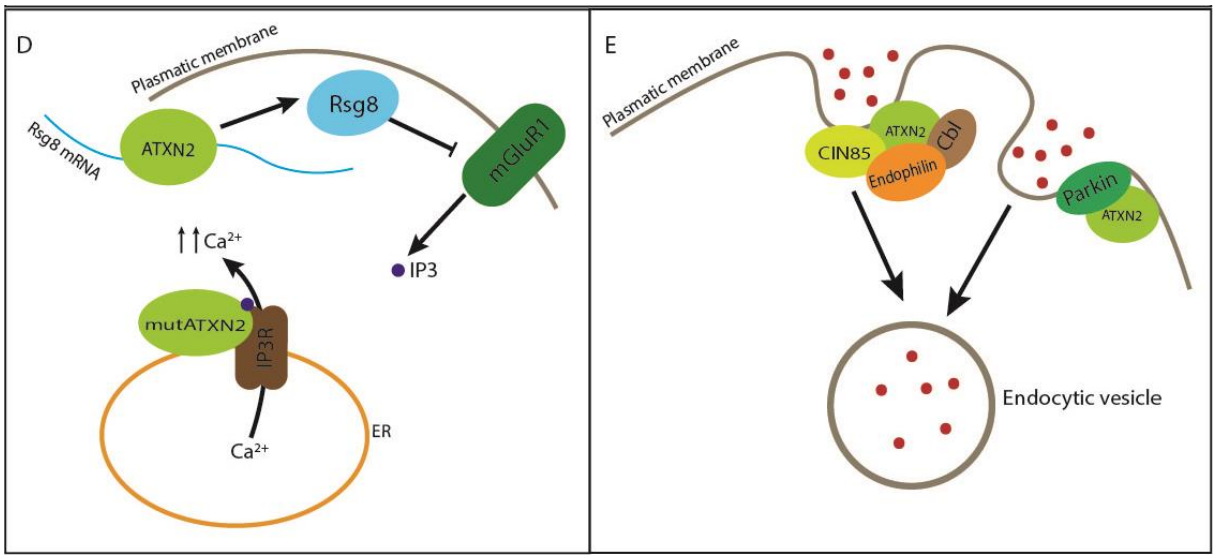

Figure 2- Different cellular processes in which ataxin-2 is implicated. A) Interaction of ataxin-2 with A2BP1 and PABP, proteins involved in mRNA alternative splicing, suggests that ataxin-2 plays a role in RNA metabolism. **B)** Association of ataxin-2 with PABP, an instrumental protein during the initiation of protein translation further hints at ataxin-2 participation in protein translation. **C)** Association of ataxin-2 with several proteins involved in SG formation, namely PABP, STAU1, TDP-43, FUS and DDX6, suggests that ataxin-2 plays a central role in regulating the assembly of SGs as well as other RNA-protein granules such as p-bodies **D)** ataxin-2 interaction with *Rsg8* mRNA and of mutant ataxin-2 with IP3R, two major players in calcium homeostasis maintenance, suggests that ataxin-2 may participate in calcium homeostasis regulation. **E)** Interaction of ataxin-2 with parkin, a modulator of endocytosis, and with an endocytosis complex containing CIN85, endophilin A3 and Cbl, propose the involvement of ataxin-2 in endocytosis.

1.1.6. Molecular bases of SCA2

PolyQ-expanded ataxin-2 with ≥ 33 glutamines displays altered biological functions, tends to aggregate and gain toxic properties, resulting in the dysfunction and death of large populations of neurons (Magaña et al 2013, Velázquez-Pérez et al 2017). Although mutant ataxin-2 displays gain-of-function features that may underly cytotoxicity, such as the ability of mutant ataxin-2 to interact with IP3R, the pathological expansion of *ATXN2* also leads to the impairment of many biological processes in which nonexpanded ataxin-2 plays a role, such as mRNA metabolism, translation and cellular trafficking. With that in mind, the pathological expansion of *ATXN2* may also be considered a loss-of-function mutation (Egorova & Bezprozvanny 2019). Overall, the pathogenic mechanisms that have been suggested to be triggered by ataxin-2 dysfunction, leading to the development of SCA2, include: protein aggregation, autophagy and mRNA processing impairment, oxidative stress, cell signaling alterations, and calcium homeostasis disruption.

Cytoplasmic and intranuclear accumulation of polyQ protein aggregates inside neurons (inclusion bodies) represents a hallmark of most polyQ disorders. Interestingly, whether these aggregates are neurotoxic or neuroprotective is still up for debate (Buijsen et al 2019). In SCA2, the formation of such aggregates may be responsible for sequestering or impairing proper trafficking of proteins essential for neuron survival, which triggers programmed cell death events, that lead to the loss of neurons from central and peripheral nervous structures (Huynh et al 2003). However, it has been proposed that the toxicity associated with ataxin-2 aggregates may result from intermediary oligomeric structures formed during the aggregation process and not specifically from the end-stage large macromolecular inclusions that are detected in the brain of SCA2 patients (Matilla-Duenas et al 2010).

Cells employ autophagy as a way of destroying polyQ-expanded mutant protein aggregates (Puorro et al 2018). Although SCA2 patients present a less prominent increase in autophagy markers in comparison to other polyQ patients, a study in ATXN2-Q127 mice, demonstrated that *ATXN2* mutation leads to abnormal autophagy (Paul et al 2018, Puorro et al 2018). These finding suggest the impairment of autophagy as a role in SCA2 pathogenesis.

Besides the possible protein-derived toxicity, natural antisense transcripts (NATs), RNAs transcribed from the template strand appear to contribute to the pathogenesis of polyQ disorders. Indeed, the *ATXN2 locus* has been demonstrated to be transcribed bidirectionally, originating an antisense transcript *ATXN2-AS*, containing CUG repeat expansions, that form RNA foci and exhibits neurotoxic activity, which may contribute to SCA2 pathology (Li et al 2016). Combining this information with the one obtained in studies performed in Huntington's disease models, that have shown discrepancy between protein aggregate formation and cell death, RNA mediated toxicity may be an additional SCA2 pathological mechanism, that acts conjunction with mutant ataxin-2 to promote cell death (Aronin et al 1999, Saudou et al 1998). Taking together these finds, with the fact that protein aggregation is less prominent in SCA2 (Puorro et al 2018), antisense transcription of *ATXN2-AS* seems to represent an additional pathogenic mechanism, that contributes for SCA2 pathogenesis.

Global transcriptome analysis of SCA2 patients indicated that ataxin-2 is involved in RNA processing of PTEN Induced Kinase 1 (PINK1), a response and quality control factor for mitochondrial stress (Sen et al 2016). Furthermore, increased levels of SOD expression accompanied by decreased levels of catalase expression were also observed in SCA2 patients. Concordantly, increased oxidative stress, abnormal mitochondrial activity, changes in the

oxidative phosphorylation system and disturbances in the antioxidant system were detected in SCA2 patient's fibroblasts. Cerebellar Purkinje cells (PCs) are highly susceptible to oxidative stress. Thus, increased oxidative stress and the resulting cell damage seem to be involved in SCA2 pathogenesis (Cornelius et al 2017).

Considering the importance cell signaling has in homeostasis maintenance, cell repair and cell communication, abnormalities in signaling mechanisms may constitute a basis for neurodegeneration (Egorova & Bezprozvanny 2019). In fact, the activity of signaling proteins has been observed to be altered in polyQ SCAs other than SCA2 (Brown et al 2018, Verbeek et al 2008), and mutant ataxin-2 can hypothetically affect cell signaling through altered RNA processing of receptors or their inhibitors, or by controlling the endocytosis of receptors from the cellular membrane. It is thus possible that aberrant cell signaling plays a role in SCA2 pathogenesis.

Fine tuning of calcium signaling is essential for normal functioning of several calcium-dependent proteins, which are implicated in synaptic transmission and calcium homeostasis, not only in PCs but in all neurons. As mentioned above, when mutated, ataxin-2 gains the ability to bind to IP3R, it leads to an IP3R hyperactivation that promotes an excessive amount of calcium ions to be pumped into the mitochondria. Calcium overload in the mitochondria leads to mitochondrial swelling, followed by outer mitochondrial membrane rupture, which culminates in apoptosis, due to the release of cytochrome C (Cyt C), a mitochondrial pro-apoptotic factor, into the cytoplasm (Egorova & Bezprozvanny 2019, Liu et al 2009).

1.1.7. Therapeutic approaches for SCA2

Currently, treatment of SCA2 is merely symptomatic. An effective disease-modifying therapy is not yet available, despite efforts made by the scientific community to find new therapeutic targets for SCA2, through the study of the physiological, functional, and biochemical events involved in SCA2 pathology. Although the molecular mechanisms involved in polyQ disorders are not fully understood, it is frequently accepted that at least part of these pathogenic mechanisms and the accompanying neurodegenerative features and symptoms are shared between this group of disorders (Nóbrega & Pereira de Almeida 2018). These similarities suggest that putative therapeutic strategies may be applied to several different polyQ diseases.

Concerning the putative therapies that are specific for SCA2, overall there are two distinct approaches behind their conception. One approach takes the known functions of ataxin-2 and its interactors as therapeutic targets, while another set targets *ATXN2* expression directly, going straight to the source of the problem behind SCA2, that is, the production of a mutant *ATXN2* transcript with a pathological CUG expansion.

Despite no curative treatment for SCA2 being available, potential treatments are currently under development and several novel therapeutic approaches remain unexplored.

1.1.7.1. Calcium stabilization

Considering that disturbed calcium signaling, and homeostasis appear to have a major role in the different types of SCAs, including SCA2, reestablishment of the physiological intracellular calcium levels represents a promising strategy to improve PC functioning and mitigate cell death. Indeed, experiments in which ER calcium channels, Ryanodine receptors (RyRs), from SCA2-58Q transgenic mice were inhibited with ryanodine or dantrolene demonstrated a normalization of the altered calcium signaling in affected PCs. Additionally, motor coordination defects were rescued, and PC loss alleviated upon long-term oral administration of dantrolene to SCA2-58Q mice (Liu et al 2009). In another study, induced pluripotent stem cells (iPSC) from SCA2 patients were differentiated into neurons and treated with dantrolene and riluzole. These two compounds also decreased intracellular calcium levels. Together, these results show that targeting and restoring calcium signaling may improve SCA2 phenotypes (Chuang et al 2019).

Pacemaker activity of cerebellar PCs regulated by small-conductance calcium-activated potassium channels (SK channel) has been demonstrated to slow down with increased levels of intracellular calcium (Meera et al 2016). With that in mind, SCA2-58Q mice were treated with NS13001, a specific positive modulator of SK2/3 channels. It was demonstrated that their abnormal firing patterns were reverted into the normal repetitive spiking, and that mice underwent a significant motor performance improvement (Kasumu et al 2012). Additional studies in SCA2-58Q mice indicated that intraperitoneal injection with chlorzoxazone (CHZ) produced a decrease in the spontaneous firing rate of cerebellar PCs, further supporting that positive SK channel modulators may present beneficial effects on ataxic PCs, by stabilizing abnormal firing rates (Egorova et al 2016).

1.1.7.2. Preventing ataxin-2 aggregation

PolyQ expanded ataxin-2 tends to form aggregates, which disrupts the native biological functions and intermolecular interactions of the protein, in addition to sequestering members of vital cell systems (Williams & Paulson 2008). Considering this, ataxin-2 aggregates represent a viable target for SCA2 therapy development, whether through the inhibition of ataxin-2 aggregation or by enhancing mutant protein clearance. No such therapy has been developed for SCA2 so far; however, several putative therapies for other polyQ diseases, including other SCAs, have shown promising results, allowing the translation of the principles behind those therapies into SCA2.

Studies performed in mouse and in fruit fly models of Huntington's disease (HD) have shown that particular peptides capable of binding to polyQ-expanded proteins are able to inhibit mutant huntingtin (Htt) aggregation and ameliorate neurodegeneration phenotypes (Chen et al 2011, Nagai et al 2003). Chaperones, such as heat shock proteins 40 (Hsp40) and 70 (Hsp70), assist the refolding and conformational maintenance of proteins (Reis et al 2017). Functional enhancement of Hsps may prevent the initiation of aggregate formation or disaggregate existing aggregating cores, thus reducing protein aggregates and the resulting toxicity (Kampinga & Bergink 2016). In fact, overexpression of chaperones presented promising results in several polyQ diseases, leading to reduced protein aggregation in cell models of SCA3 (Chai et al 1999) and SCA1 (Cummings et al 1998). Moreover, results in SCA1 (Warrick et al 1999) and SCA3 (Bilen & Bonini 2007) fly models demonstrated that chaperone overexpression suppressed neurodegeneration; additionally, in SCA1 mouse models (Cummings et al 2001), this type of strategy improved motor function.

Although protein aggregation in SCA2 is not as pronounced as in other PolyQ diseases, ataxin-2 appears to be implicated in autophagy, considering that *ATXN2* mutation results in an impairment of autophagy. Importantly, stimulating the degradation of mutant proteins may normalize cellular systems disrupted by polyQ toxicity (Huynh et al 2000, Matos et al 2018). With the aim of enhancing autophagy, SCA17 cell and mouse models were treated with trehalose, an autophagy activator through mTOR signaling. As a result, significant reduction of protein aggregates and improvement of ataxic behavior and motor coordination were observed (Chen et al 2015). Furthermore, trehalose analogues applied in SCA3 cell models presented similar results (Lin et al 2016). In another approach to promote autophagy**,** neuronal cultures and a SCA3 rat model were transfected with a lentiviral vector expressing Beclin-1, a protein

important in initiating autophagy. This approach, which stimulated autophagic influx, led to mutant ataxin-3 clearance and neuroprotective effects (Nascimento-Ferreira et al 2011).

1.1.7.3. Regulating ataxin-2 pos-translational modifications

Post-translational modifications such as phosphorylation and proteolytic cleavage have a major role in regulating several protein properties, influencing protein localization, stability, function, aggregation, and interaction with other molecules. Modulation of the polyQ disease-causing proteins by inducing alterations in their post-translational modifications has the potential to affect the underlying pathogenic mechanisms in a disease-modifying manner, conferring therapeutic benefits to the affected cells (Wan et al 2018).

Notably, studies performed on COS-7 cells expressing both normal and mutant ataxin-2 fragments have demonstrated that cyclin-dependent kinase 5 (Cdk 5)-mediated phosphorylation of ataxin-2 has an important role in SCA2 pathology. Remarkably, Cdk-5-mediated phosphorylation was shown to control the abundance of both normal and mutant ataxin-2 by inducing their degradation. The same study also demonstrated that Cdk5 activity has therapeutic significance, considering that toxic ataxin-2 underwent proteasomal degradation after Cdk5-mediated phosphorylation (Asada et al 2014).

Evidence in several polyQ disorders supports the notion that generation of intracellular toxic protein fragments is an important part of disease pathogenesis. PolyQ sequence-containing fragments generated through proteolytic cleavage have been described to possess greater cytotoxic properties comparing to the corresponding intact full-length proteins. Thus, polyQ protein fragments are considered by many as the actual cause behind protein toxicity and consequent cell death (Buijsen et al 2019, Matos et al 2018). Huntingtin and ataxin-3 fragments have been suggested to have their origin in proteolytic fragmentation promoted by proteases such as calpains and caspases (Gafni & Ellerby 2002, Haacke et al 2007, Kim et al 2001, Simoes et al 2012). Notably, inhibition of calpains in the brain of a SCA3 mice model by AAV-mediated overexpression of calpastatin, a calpain inhibitor, resulted in the decrease of aggregation and neurodegeneration (Simoes et al 2012).

Similarly to huntingtin and ataxin-3, ataxin-2 may stimulate the activation of caspases and be cleaved by them, leading to the formation of truncated proteins with enhanced toxicity (Huynh et al 2003). In fact, ataxin-2 presents a putative caspase-3 cleavage site, that was

predicted to produce a fragment with the same molecular weight as the fragments of ataxin-2 that have been experimentally detected. Thus, caspase-3 was suggested to promote proteolytic fragmentation of ataxin-2 (Shibata et al 2000). A therapy inhibiting caspase 3, and halting ataxin-2 proteolytic fragmentation, could be potentially beneficial for SCA2, by reducing the formation of truncated forms of ataxin-2 and their resulting toxicity.

1.1.7.4. Multipotent stromal fibroblast-like cells transplantation

Multipotent stromal fibroblast-like cells (MSCs) can differentiate into several different cell types and exhibit immune-modulatory functions, which include the release of neurotrophic factors (Chang et al 2011). Upon getting transplanted into the cerebellum of Lurcher mice, a model presenting olivocerebellar degeneration, MSCs were shown to migrate towards the Purkinje cell layer and express neurotropic factors implicated in PC survival, such as brain-derived neurotropic factor (BDNF), neurotrophin-3 and glial cell-derived neurotrophic factor (GDNF) (Jones et al 2010).

When delivered intravenously to transgenic SCA2 mice, human MSCs delayed the onset of motor coordination dysfunction and improved motor function by preventing PC loss (Chang et al 2011). The feasibility of this kind of therapeutic approach was proven by a clinical study in which six SCA3 patients were injected with MSCs. Notably, most injected patients displayed significant improvements in cerebral and cerebellar activity, as well as enhanced performance in sensory organization tests. Additionally, the results obtained in this study also suggest safety and tolerability of the treatment (Tsai et al 2017).

1.1.7.5. Targeting the root cause of SCA2 – silencing ataxin-2

In the context of polyQ disorders, perhaps the most straightforward approach to therapy would be to silence the causative mutant polyQ proteins. Since they are the root cause of these diseases, inhibiting expanded polyQ protein expression would admittedly prevent disease progression at the earliest step possible. Moreover, the fact that polyQ diseases have their underlying cause in singular genetic factors makes them ideal targets for gene silencing approaches.

Gene silencing in SCA2 has an increased interest. Since mutant ataxin-2 may disrupt several cellular pathways, targeting separate pathways or proteins mediating ataxin-2-derived toxicity would admittedly affect only those particular pathways, and not the other cellular processes in which the pathogenic protein is implicated. Contrariwise, *ATXN2* silencing allows for a reduction in mutant polyQ expanded protein levels, which will ultimately affect all the downstream pathological pathways, ideally hindering disease onset and/or progression (Buijsen et al 2019, Gatchel & Zoghbi 2005).

In the context of the treatment of human diseases, silencing of a pathogenic gene may be achieved through a set of different strategies overall contained in the wider field of gene therapy. The following sections discuss these innovative approaches.

1.2. Gene therapy

A large portion of human diseases is caused by the dysfunction of genes, resulting from genetic mutations, deletions, or disruptions. Gene therapy aims at counteracting harmful gene dysfunctions through the delivery of nucleic acids, for example by supplementing a functional copy of a defective gene. Its main goal is to achieve a durable expression of the therapeutic gene, sufficient to mitigate symptoms or cure the disease, while causing minimal adverse events. Contrary to protein and peptide-based drugs, gene-based therapies generally target the underlying cause behind a disease, directly. Certain strategies of gene delivery allow for durable or permanent expression of a therapeutic gene coding for the desired protein with a single application, contrary to protein-based drugs, which require multiple administrations (Dunbar et al 2018, High & Roncarolo 2019).

Although gene therapy was formerly viewed solely as a treatment for inherited disorders, it is currently regarded as being applicable to acquired conditions, most frequently cancer. In this case, rather than delivering a gene that substitutes a dysfunctional allele, a "suicide" gene that promotes the death of tumor cells may be delivered. For example one of the most used suicide genes is the gene coding for thymidine kinase, a key protein in DNA synthesis and cell division (Dunbar et al 2018, Düzgüneş et al 2018).

1.2.1. Gene therapy delivery systems

The ideal vector used in the delivery of nucleic acids should have a high transduction efficiency, should be sufficiently spacious in order to allow carrying a large payload and should also possess the ability to provide stable and long-lasting gene expression. Additionally, it should transduce only specific cells, and avoid complications due to random insertion of the gene of interest or immune responses from the host. Furthermore, the ability to be manufactured on a large scale should preferably be present, in order to allow for a scale up in production and thus meet up with the high quantities required to deliver efficiently the therapy to a fully grown human (Goswami et al 2019).

A wide variety of physical and chemical delivery methods have been developed and applied for administering gene-based therapies. These include gene guns, microneedles, electroporation, sonoporation and magnetofection procedures, calcium phosphate-mediated transfection, lipidic vesicles, gold nanoparticles and cationic polymers (Matos et al 2018, Sum et al 2018). Considered to be Nature´s genetic engineers, viruses have the intrinsic ability to infect most organisms, such as bacteria, plants, humans, and animals, and deliver functional nucleic acids. None of the vectors referred above is more efficient than viral vectors in the delivery of therapeutic genetic material into the intended cells. This makes viral vectors the main delivery strategy used to deliver gene-based therapies in preclinical and clinical investigations (Goswami et al 2019). There is an abundance of different delivery methods available, each with its own pros and cons; it is very important to take into account the objective of the particular therapeutic approach when selecting the respective delivery method, in order to choose the one that is the most suitable to deliver the payload of interest to the target cells.

Despite the great variety of vectors available to choose from, two groups take the central stage: the retroviral and the adeno-associated viral (AAV) vectors. Unlike all other RNA viruses that are employed, retroviral vectors, with 9-12 kb of package size, are capable of undergoing reverse transcription and integrate DNA at a desired genetic *locus*, allowing the therapeutic gene to be passed to every daughter cell (Dunbar et al 2018). In contrast to retroviruses, AVV vectors are predominantly nonintegrating and only allow for a ~5 kb packaging capacity, but grant the ability to delivery genes to both mitotic and post-mitotic cells, while providing a durable expression of the therapeutic gene to the cell (High & Roncarolo 2019). The different naturally occurring AVV serotypes and chimeric AVVs feature less risk of immune response

than other viral vectors; each different serotype possesses tropism to specific tissues, enabling for a much more selective delivery of the intended payload (Dunbar et al 2018).

Nowadays, gene therapy goes hand-in-hand with cell-based therapy. Stem cells can be extracted from a patient and transduced *ex vivo* with the therapeutic gene while in culture, which will allow the transduced stem cells to become healthy. The transduced stem cells are then returned to the patient by transplantation, possibly after being stimulated to differentiate into relevant cell types. On the other hand, *in vivo* delivery of gene therapies does not require the cells to be extracted from the patient; instead, the delivery of the therapeutic gene to long-lived postmitotic cells resembles the administration of other types of pharmacological agents, as the cells are directly transduced in the body of the patient. *In vivo* cell transduction can take advantage of the tropism of certain vectors that enables them to find the intended target, when delivered systemically (High & Roncarolo 2019, Matos et al 2018).

1.2.2. The promise and challenges of gene therapy

In the infancy of gene therapy, the scientific community had high hopes towards such promising and novel approach to disease treatment. However, after studies showed inconclusive clinical benefits and unexpected deleterious outcomes, such as insertional genotoxicity, a decrease in the *momentum* of gene therapy was expected (Dunbar et al 2018). Despite the unexpected results, gene therapy experienced a rise in popularity until 1999, when 18-year old Jesse Gelsinger, who took part in a gene therapy clinical trial, died due to multiorgan failure resulting from immune response to a very high dose of adenovirus administrated four days earlier (Stolberg 1999). This tragic event created awareness to the risk of using viral vectors and made researchers focus their efforts in expanding the safety of this type of therapy. Over the past decades, many of the concerns underlying gene therapy have been addressed through safety modifications, enhancements in gene transfer efficiency and improvement of delivery methods. The increased safety profile and the subsequently boosted trust in gene therapy culminated in the approval of diverse gene and genetically modified cell-based therapies by the regulator entities of United States and Europe, in recent years (Dunbar et al 2018).

Regardless of the risks still associated with gene therapy, the collaboration between the academia, the industry and the regulators has made it safer than ever to perform such therapeutic approaches (Dunbar et al 2018), and as a result the number of approved clinical trials continues

to increase (clinicaltrials.gov). Challenges still exist, and the possibility of developing uncontrolled genetic modifications undoubtedly still poses a risk to the individual. Even more severe consequences may arise if unwanted genetic modifications are passed down to the offspring (Wirth et al 2013).

It is important to stress that, although viral vectors are associated with many of the former risks of gene therapy, this type of delivery system played a central role in the development of gene therapy, having been employed in the initial studies that established gene addition, that is, adding a wild-type copy of a recessive gene, as a viable therapeutic approach to the treatment of recessive genetic diseases (Goswami et al 2019, High & Roncarolo 2019). To this day, viral vectors are still the delivery method of choice and see extensive use in other types of applications. Viral vectors are combined with versatile "effector" tools, including RNAi technology, which enacts gene silencing (Nóbrega et al 2013), and gene editors, that promote gene editing/silencing (Ran et al 2015) (both discussed in section 1.2.3 and 2.4, respectively).

1.2.3. RNAi and ASOs

The understanding of gene regulation in eukaryotic cells was transformed when, in 1998, double-stranded RNAs (dsRNA) were found to regulate many cellular processes through post-transcriptional gene silencing in *Caenorhabditis elegans* (Fire et al 1998). The description of this processed, named RNA interference (RNAi), would later grant a Nobel Prize in Physiology or Medicine to Andrew Fire and Craig C. Mello, in 2006 (https://www.nobelprize.org/).

Physiological RNAi is driven by microRNAs (miRNAs), endogenous non-coding RNAs comprised of approximately 22 nucleotides. These non-coding RNAs have been found to regulate different cellular processes, from cell proliferation, and development, to cell death (Bartel 2004). Taking advantage of the discovery of this versatile tool for regulating target genes, researchers started to synthesize artificial RNAi molecules constituted by 21 to 23 nucleotides, with the aim of inhibiting genes of interest and observing the outcome, as part of fundamental research efforts. This opened the door for designing oligonucleotides capable of silencing a gene responsible for a particular pathology, enabling the treatment of dominant genetic diseases (Setten et al 2019). These artificial RNA duplexes capable of triggering RNAi

may be delivered in different ways. They can be introduced into the cell as small interfering RNAs (siRNAs), similarly to a protein-based therapy, or via plasmids and viral vectors, which incorporate into the genome and get endogenously expressed in the form of short hairpin RNAs (shRNAs), which, contrary to siRNAs, result in durable gene silencing (Han 2018). Experimentally, the mechanism of RNAi can thus be triggered by different RNA molecules: miRNAs, siRNAs and shRNAs (Matos et al 2018).

In the case of nucleus-bound RNA molecules such as miRNA and shRNA, the RNAi pathway starts with the expression of those RNAs transcripts, containing hairpin structures, and their processing by Drosha. The processed RNA molecules are then transported to the cytoplasm via exportin-5, a nuclear membrane protein. Once in the cytoplasm the RNAi pathway is common to the nuclear and cytoplasmic RNAs, such as siRNAs. Upon entering the cytoplasm, RNAi molecules are processed by Dicer, giving rise to a mature RNAi molecule. The antisense strand from the mature RNAi molecule is then load onto a protein complex named RNA-induced silencing complex (RISC), which will recognize a specific complementary mRNA. Upon biding, two silencing pathways can occur: a RNAse-mediated degradation when the target sequence mRNA is 100% complementary to the RNAi molecule, which results in digestion of that target sequence, or repression of mRNA translation when complementarity is partial (Bobbin & Rossi 2016, Matos et al 2018).

Even after discarding delivery vector-related toxicity, therapeutic approaches based on RNAi present critical challenges. Nonspecific toxicity may arise due to innate immunogenic reactions to foreign double-stranded RNAs introduced into the cell or due to off-target RNAi activity, caused by complete or incomplete matching of the designed oligonucleotides to non-target mRNAs. However, with the development of bioinformatic tools capable of designing optimal sequences and predicting their putative binding targets and offsite effects, silencing sequences can be designed to present maximum therapeutic effect (Setten et al 2019). Additionally**,** a refining process where a silencing sequence is designed in order to target a SNP in a mutant messenger RNA adds to RNAi molecules the ability to discriminate between wild-type and mutant forms of mRNAs, as well as granting an added level of specificity that reduces possible off-targets (Nóbrega et al 2013).

Many RNA interference-based therapeutic strategies have been developed for several SCAs and other polyQ diseases, as reviewed in (Buijsen et al 2019, Matos et al 2018) respectively. Some of these approaches aimed at treating SCA6, SCA7 and SCA1 disease

models with microRNAs. In the case of SCA6, this resulted in an alleviation of motor deficits and Purkinje cell degeneration (Miyazaki et al 2016), while in SCA7 a reduction of ataxin-7 levels of up to 80% was observed (Tan et al 2014). Finally, in SCA1, both *ATXN1* mRNA and ataxin-1 protein levels were significantly reduced, and an improvement in motor performance and neuropathology was observed (Keiser et al 2014). Considering the promising results observed in these other SCAs, targeting *ATXN2* mRNA using iRNA may represent a viable option to act upstream of the SCA2 pathological pathway.

Antisense oligonucleotides (ASOs) can be used as an alternative to RNAi technology. ASOs are synthetic oligonucleotide mimetics, which are able to bind specific mRNA molecules and block their expression (Matos et al 2018). ASOs-based therapy for SCA2 represents one of the most promising therapeutic approaches currently under development for the disease. This therapy consists in targeting the *ATXN2* mRNA with a complementary sequence of modified oligonucleotides, used to suppress its translation and thus reduce the level of mutant ataxin-2 expression. Upon choosing the most suitable oligonucleotide to target *ATXN2* mRNA through a screening process, a particular molecule (ASO7) was delivered to SCA2-127Q and BAC-72Q transgenic mice models via intracerebroventricular injections. Human *ATXN2* expression in the mouse models underwent a reduction of 75% and there was a significant improvement in motor function. Slow firing frequency of PCs was also restored in both models. Notably, protein levels of several SCA2-related proteins expressed in PCs, namely Rgs8, were normalized upon treatment with ASO7 (Scoles et al 2017).

1.3. Gene editing

Conventional gene therapy methods are limited to addition of exogenous nucleotide sequences into the genome, whereas gene editing methods allow for precise and targetable modification of genome sequences, enabling the knockout of endogenous genes or the correction of existing defects (Kc & Steer 2019, Maeder & Gersbach 2016). Furthermore, in contrast with RNAi strategies, which reduce but do not completely eliminate gene expression, gene editing can be used to completely block gene expression. This complete disruption can be achieved by altering the genome so as to inhibit binding of the transcription machinery to the DNA, or by introducing modification that alter the translation process of its correspondent mRNA (Bak et al 2018).

Induction of a DNA double-strand break (DSB) within a region sharing homology with another DNA sequence was observed to greatly enhance the frequency of homologous recombination between the two DNA sequences (Fernandez et al 2017). The discovery that targeted DSBs could be used to stimulate the endogenous DNA repair mechanisms would become foundational to the field of gene editing, which relies on the fact that, if the appropriate conditions are created, those DNA repair mechanisms can be harnessed to execute deliberate alterations in the genome (Fernandez et al 2017, Maeder & Gersbach 2016). Typically, there are two major DNA repair pathways responsible for repairing DNA breaks: nonhomologous end-joining (NHEJ), which occurs in the absence of template DNA molecules, and homology-directed repair (HDR), which occurs when they are a present.

NHEJ is active during all stages of the cell cycle and re-ligates the cleaved end of the DNA directly, without the need for a repair template. This repair pathway is prone to errors and often introduces small insertions and/or deletions (indels) at the repair site. With this in mind, inducing site-specific DSBs can be used to stimulate NHEJ and take advantage of indels to disrupt target genes in a vast array of organisms and cell types. This occurs when the indels shifts the reading frame of the targeted gene or introduces premature STOP codons (Bak et al 2018, Fernandez et al 2017, Maeder & Gersbach 2016).

Contrary to NHEJ, the HDR pathway is mainly active during the S and G2 phases of the cell cycle and utilizes the endogenous DNA of a sister chromatid as a repair template. HDR relies on strand invasion of the homology template by the broken ends of the cleaved strand, and the subsequent DNA synthesis of a complementary strand based on the template. Lastly, there is separation between both DNA helixes and the newly synthesized sequence (or, alternatively, the template itself) will end up being integrated into the previously broken DNA molecule. Exogenous donor DNA sequences may also be used as a template for DSBs repair in case those sequences contain segments that are homologous to the vicinity of the DSB site. In the presence of this type of template, inducing a DSB of a DNA molecule enables HDR to precisely insert an exogenous DNA sequence at the specific genomic *locus* that was targeted**.** Competition between the two DNA repair pathways - NHEJ and HDR - tends to reduce template addition efficiency, due to cells preferentially using NHEJ to repair DSB (Bak et al 2018, Fernandez et al 2017, Maeder & Gersbach 2016, Urnov et al 2010).

Targeted gene editing relying on these DNA repair mechanisms presupposes the ability to produce DSBs at particular genomic sites, i. e, at the vicinity of the sites intended to be altered. DSBs can be artificially introduced into the genome through the action of specific endonucleases, which can be engineered and repurposed for different applications (Maeder & Gersbach 2016). Currently, endonucleases used in gene editing are grouped into four main classes, which will be discussed in the following section. Efficient application of gene editing requires many different aspects to be considered, and selection of the particular endonucleases used to introduce the DSBs is a critical step of the process. The most significant idea is the necessity to target DNA sequences with high efficiency and specificity within the genome, ensuring the absence of "off-targets" effects that may cause unintended complications. With the objective of addressing this need, the ability of several proteins to bind to the DNA was exploited, either by reengineering their DNA-binding domains or fusing multiple DNA-binding domains together (Jinek et al 2012).

1.3.1. Meganucleases

Meganucleases, also known as homing endonucleases, are enzymes encoded within self-splicing RNA introns or self-splicing protein introns (inteins), that have the ability to function autonomously, recognizing and cleaving long specific DNA sequences with 12 to 40 bp (Seligman et al 2002). Meganucleases are divided into five families according to their sequence and structural motifs, with the most well studied and widely used family being the LAGLIDADG family, which includes the commonly used enzymes I-*Cre*-I and I-SceI (Maeder & Gersbach 2016, Seligman et al 2002, Silva et al 2011)**.**

Studies made using I-*Sce*I laid the foundation for modern genome engineering strategies, elucidating DSB repair mechanisms and limits to the nuclease-based homologous recombination methodology (Silva et al 2011). I-*Sce*I-induced DSBs proven to be efficient in promoting homologous recombination with exogenous DNA molecules carrying regions that were homologous to the I-*Sce*I restriction site in mammalian cells (Fernandez et al 2017). This type of observations suggested the possibility of treating and array of dominant monogenic diseases through the correction of the mutations responsible, by substituting the disease-associated gene segment with a "healthy" copy of this segment, in affected cells.

Major limitations to strategies using meganucleases include the need for a compatible homing nuclease cleavage sequence in the host genome or the prior introduction of homing endonucleases natural cleave site in the *locus* of interest (Arnould et al 2007). To combat these limitations, meganucleases have had their DNA-binding domain re-engineered through different strategies, such as swapping domains between I-*Cre*I and I-*Dmon*I, leading to the generation of new hybrid meganucleases. Alternatively, altering the DNA recognition properties of I-*CreI* and grouping it in arrays, allowed for the development of an enzyme that can recognize and cleave a chosen sequence (Seligman et al 2002, Silva et al 2011).

1.3.2. Zinc finger nucleases (ZFNs)

Kim and collaborators were the first research group to describe the potential of zinc finger proteins (ZFPs) in gene editing and to design a chimeric zinc finger nuclease (ZFN) (Kim et al 1996). Mainly two domains form a zinc finger nuclease: a flexible and robust DNA-binding domain from zinc finger proteins, and a catalytic domain from the FokI restriction enzyme (Kim et al 1996, Urnov et al 2010).

FokI is the most widely used restriction enzyme in the design of chimeric nucleases, not only of ZFNs, but also TALENs (discussed below in section 1.3.3.). This restriction enzyme was isolated for the first time from the bacteria *Flavobacterium okeanokoites* (Sugisaki & Kanazawa 1981). Structurally, this enzyme is comprised by a N-terminal DNA-binding and a C-terminal DNA-cleavage domain, that cleaves the DNA double strand non-specifically, at 9 and 13 nucleotides downstream of the recognition site. In the absence of divalent metals and without DNA binding, the FokI monomer is sequestered and maintains an idle state. Upon binding of two FokI holoenzymes to the DNA recognition site, in the existence of divalent metals, the monomers dimerize, allowing for catalytic activity and subsequence cleavage of the DNA double-strand (Li et al 2011).

The ZFN DNA-binding domains take advantage of the DNA recognition properties from Cys2-His2 ZFPs, a class of ZFPs containing the most common DNA-binding domain motif in eukaryotes, as well as the most prevalent in the human genome (Segal et al 2003). The DNA recognition site of ZFP is organized in several tandem arrays of DNA-binding units, each of which is being capable of recognizing a three-nucleotide subsite, also called triplet (Kim et al 1996, Rebar & Pabo 1994, Segal et al 2003). ZFPs offer an attractive framework for the

design of chimeric nucleases, due to their modular structure, formed by these independent tandem DNA-binding domains. ZFN protein engineering is achieved through the use of recombinant DNA technology such as modular assembly. Modular assembly links together several DNA-binding domains, in order to create a custom scaffold with specificity for a particular DNA sequence of interest, that can posteriorly be fused to an endonuclease - usually FokI, as mentioned (Kim et al 2009, Kim et al 1996).

ZFNs were critical for the empowerment of gene editing, increasing the amount of nucleotide sequences that could be recognized through the employment of a nuclease that is much more versatile than its predecessors – the meganucleases. Since DNA-binding domains (fingers) targeting most triplet sequences have been developed, fusing them together in a multi-finger peptide enables to creation of customizable nucleases with increased specificity and able to target DNA sequences that occur more rarely in the genome. Further accuracy is reached due to the catalytical requirements of the FokI enzyme, which must dimerize in order to promote DNA DSB. The use of ZFN dimers that bind to the anti-parallel DNA chains and promote DSB catalyzed by FokI reduces the probability of off-target effects, since the occurrence of the two target sequences in an antiparallel manner, outside the target *locus*, is highly unlikely (Miller et al 1985, Rebar & Pabo 1994, Urnov et al 2010).

Even though ZFN technology in fundamental and applied scientific research seemed promising, widespread adoption of this nuclease was halted by barriers such as the intensive labor and lengthy time associated with the design and engineering of custom zinc fingers with high affinity and specificity for a DNA target site of interest (Li et al 2011). Furthermore, assembled ZFNs, validated *in vitro*, showed high rates of failure *in vivo*, failing to drive gene editing at endogenous loci. The inability for driving gene edition at endogenous loci might be due to the complexity of the genome, which can accommodate sequences that identical or highly related to the target site. This, in addition to the variable state of chromatin condensation, may account for the failure of a precise DNA biding and cleavage process. Improper assembly of ZFP domains that recognize four nucleotides instead of three into a ZFNs modulus can also limit its ability to specifically bind to the target site (Segal et al 2003). Advances in bioinformatics and ZFPs assembly were made in order to address these limitations, but the emergence of TALENs would cause a shift in efforts and dampen future progress.

1.3.3. TALENs

Similarly to ZFNs, transcription activator-like effector nucleases (TALENs) consist in the fusion of a DNA-binding protein domain, in this case a transcription activator-like effector (TALE), with the restriction enzyme FokI. Characteristically, the FokI enzyme requires the dimerization of two TALENs in an antiparallel fashion to cleave the DNA region flanked by both TALE DNA-recognition sites (Beurdeley et al 2013).

TALEs, first discovered in the *Xanthomonas* genus, were described as a member of a large group of proteins highly conserved in bacteria that are responsible for causing severe diseases on many crop plants. Furthermore, TALEs were demonstrated to be a key factor in the bacteria's virulence; once in the cell nucleus, TALEs bind to effector-specific DNA sequences and activate expression of specific target genes, promoting an increase in the plant susceptibility to bacterial multiplication (Beurdeley et al 2013, Boch et al 2009, Bonas et al 1989, Christian et al 2010).

TALEs are constituted by 3 domains, including a central domain of tandem repeats, a nuclear localization signal (NLS) and an acidic transcriptional activation domain (AD) (Boch et al 2009). The central tandem repeats, consisting usually of 34-amino acids repeats, are responsible for recognizing specific DNA sequences. Each repeat is nearly identical apart from a two amino acid polymorphism in the positions 12 and 13, a motif that is called repeat-variable di-residue (RVD). The DNA binding specificity is based on the RVDs present in a repeats, since each RVD is responsible for the ability of the repeat to bind directly to a particular base pair. Consequently, a particular tandem array of repeats is able to bind a designated consecutive DNA sequence (Beurdeley et al 2013, Boch et al 2009, Li et al 2011, Moscou & Bogdanove 2009).

TALEs enable a versatile targeting of specific DNA sequences, by engineering the amino acids in the RVDs using the cipher that describes the correspondence between one RVD and one nucleotide (Boch et al 2009). Moreover, most naturally occurring TALE proteins recognize lengthy DNA target sites, since the central tandem repeats enable the binding of the protein to a maximum of 34 nucleotides. Using the modular assembly method previously utilized to engineer ZFNs allowed the efficient construction of TALE DNA-binding scaffolds

(TDBS) that could be fused to different enzymes, most commonly to the FokI restriction enzyme (Beurdeley et al 2013, Li et al 2011).

Although promising, TALEN monomers are roughly three time larger than other engineered nucleases, which constitutes a significant disadvantage. Whether in DNA, RNA or protein format, delivery of such large payload presents many challenges (Beurdeley et al 2013).

1.3.4. CRISPR-based gene editing tools

While the previous gene editing platforms solely rely on protein constructs that bind directly to their DNA targets, the CRISPR/Cas system consists of two distinct components - an endonuclease, usually the Cas9 protein from *Streptococcus pyogenes*, and a guide RNA molecule, that combines with Cas9 to form a ribonucleoprotein complex (Doudna & Charpentier 2014). The guide RNA directs the complex to its DNA target by binding to the particular nucleotide sequence through classic Watson and Crick base-pairing. Contrary to singular ZFNs or TALENs, Cas9 is then able to cleave both strands of the DNA double helix on its own, requiring no dimerization (Hsu et al 2014, Mali et al 2013a)

The fact that retargeting the CRISPR/Cas system requires only the design of a new guide RNA sequence contrasts with the lengthy molecular cloning protocols that reengineering chimeric endonucleases entails. This represents an unprecedented advantage over the other gene editing platforms and allows CRISPR/Cas-based gene editing to be performed with relative ease, in a variety of living systems. This and other advantages led to the rapid adoption of the CRISPR/Cas system as the gene editing tool of choice in laboratories all over the world, and therapeutic applications based on this platform are already being tested in human subjects (Adli 2018, Xu et al 2019).

1.3.4.1. From peculiar prokaryotic DNA loci to gene editing tools

Clustered regularly interspaced short palindromic repeat (CRISPR) DNA sequences were initially described in a study by Ishino and colleagues in *Escherichia coli* (Ishino et al 1987), and later in *Haloferax mediterranei*, by Francisco Mojica and collaborators (Mojica et al 1993). Notably, these DNA sequences consisted in a series of tandem repeats varying in size from 30 to 34 bp, separated from each other by 35-39 bp long non-repeating sequences

(spacers). This pattern was unlike typical tandem repeats in the genome. More than a decade went by until researchers would be able to recognize the origin and function of these sequences, through *in silico* analysis (Adli 2018).

The rise of computational analysis as a method to break down genomes led to different researchers noticing key features of CRISPR loci. It began with Ruud Jansen and his colleagues identifying multiple well-conserved genes adjacent to CRISPR elements, a term this group was the first to instate. These four well-conserved genes, called CRISPR-associated (Cas) genes, were designated as *cas1* through *cas4*. The *cas1* gene was found in all species containing the CRISPR loci, whereas the other three *cas* genes, although present in most CRISPR-containing species, were not identified in all analyzed genomes. Despite no conclusive function being attributed to *cas*-encoded proteins, by analyzing the amino acid sequences of Cas3 and Cas4, the group was able to predict similarity to the superfamily 2 of the helicases and RevB exonucleases, respectively (Jansen et al 2002). Apart from the CRISPR loci previously described multiple distinct subtypes of CRISPR loci have since been characterized, demonstrating great differences in *cas* genes content between approximately related strains and species of bacteria (Haft et al 2005).

In 2005, studies from Francisco Mojica and Alexander Bolotin, which compared CRISPR sequences with natural genetic elements, came to the conclusion that, contrary to what was described in early CRISPR related studies, CRISPR spacers sequences have their origin in exogenous genetic material, namely from bacteriophages and mobile genetic elements (Bolotin et al 2005, Mojica et al 2005). Most genes detected in the studies by Mojica's and Bolotin's groups were shown to be directly involved with the infectivity of phages towards bacteria, therefore suggesting that CRISPR could be involved in conferring immunity against foreign DNA. However, the scarce knowledge at that time limited the delineation of any particular action mechanism. It was not until 2007 when Rodolph Barrangou and his team proved that CRISPR was indeed responsible for an adaptive immune system present in prokaryotes (Barrangou et al 2007). In their study, they demonstrated that bacteria are subjected to dynamic and rapid evolutionary changes upon exposure to bacteriophages, by being able to integrate the genomic material derived from infectant phages into the CRISPR *loci*. The integrated sequences constitute the spacer sequences. Furthermore, after altering the CRISPR *locus* in bacteria by adding or deleting spacers and subsequently exposing them to a certain strain of bacteriophages, they showed that spacer content was able to provide phage resistance, even though spacers alone did not provide resistance. This resistance is in fact provided in conjunction with the Cas

enzymatic machinery, as observed after inactivating *cas5* and *cas7* genes. Just a year later, in 2008, Stan Brouns and colleagues showed that the transcribed regions of CRISPR undergoes a maturation process, through the cleavage of pre-CRISPR RNA (pre-crRNA) to CRISPR RNA by action of a Cas protein complex. The Cas proteins and the CRISPR RNA constitute the molecular basis for the efficient antiviral defense in prokaryotes (Brouns et al 2008).

Different evolutionary paths taken by prokaryotes led to the development of different adaptive immune systems, in response to the exposure to distinct environmental mobile genetic elements. As a consequence, CRISPR/Cas systems differ between species of prokaryotes. To address this diversity, CRISPR/Cas systems were divided into two classes, each one further comprising multiple CRISPR types that differ in the mechanisms used to form cas protein complexes, the proteins that constitute them and how they provide immunity. In summary, Class 1, commonly found in Archaea, contains the type I and type III CRISPR systems, while Class 2, which has been found exclusively in Bacteria, is comprised of the type II, IV, V and VI systems (Koonin et al 2017). Cas proteins are a highly diverse group, functioning as nucleases or RNA-binding proteins. Their corresponding genes occur in certain CRISPR *loci* and not in others; when present, these genes can be differentially arranged, even among CRISPR *loci* belonging to the same type. Despite this diversity, Cas1 and Cas2 proteins, which are involved in the insertion of new spacers, are virtually universal among CRISPR systems, although the adaptation process varies greatly between the various system types, depending on the specific *cas* genes expressed along with *cas1* and *cas2* (Makarova et al 2011).

The CRISPR array, which contains the spacers, the repeats and which codifies the CRISPR RNA (crRNA) is also present in all the different prokaryote adaptive immune system (Adli 2018, Rath et al 2015). The crRNA is responsible for guiding the protein (or protein complex) that cleaves and degrades the exogenous DNA, binding to a complementary sequence of 20 nucleotides present in the foreign element. In order to be targeted, the genetic sequences of the exogenous DNA that are complementary to the crRNA – named protospacers – must be followed downstream by a protospacer adjacent motif **(**PAM), which is necessary for complementary sequence recognition (Jinek et al 2012). This nucleotide motif varies between different CRISPR systems, and even among systems coexisting in the same organism.

Some of molecular agents – proteins and RNA molecules – from several different naturally-occurring CRISPR systems have been modified and adapted to serve as biotechnological effectors of gene editing. As mentioned above, the CRISPR system that has

been most commonly employed and that has gained the most traction in the scientific community is a type II CRISPR system from *S. pyogenes*. Composed by three distinct regions, the corresponding native CRISPR *locus* comprises a trans-activating crRNA (tracrRNA) gene, the *cas* genes (*cas9*, *cas1*, *cas2* and *csn2/cas4*) and the CRISPR array (leader, repeats and spacers), which when transcribed gives origin to the pre-CRISPR array, that is later processed into several crRNAs that are different depending on the corresponding spacer. Although the precise function of the tracrRNA and some *cas* genes remain to be determined, it is known that Cas1 and Cas2 complex assembly is essential for the acquisition and incorporation of foreign DNA spacers into the CRISPR array, and that, uniquely to type II systems, the Csn2 protein, which binds to DNA, is involved in the prokaryote adaptive immune through the formation of the Cas1-Cas2-Csn2 complex. Importantly Cas9 has the ability to induce DSB of exogenous DNA as a single protein, contrasting with CRISPR systems in which several proteins have to come together as a complex to cleave foreign DNA (Ka et al 2016, Nunez et al 2014). This simplicity constituted an advantage that made this particular CRISPR system easily amenable to biotechnological repurposing.

As with the other CRISPR/Cas-mediated immune response to foreign genetic elements, immunity mediated by the type II CRISPR system of *S. pyogenes* consists of three stages – adaptation, expression, and interference. During the adaptation stage, a sequence of foreign DNA is cleaved as defined by the presence of the appropriate PAM in the invading DNA. In the case of the type II CRISPR system of *S. pyogenes*, the PAM consists of a NGG consensus sequence, comprised of a variable nucleotide pair followed by two G:C base pairs (Jinek et al 2012). Upon cleavage, the protein complex Cas1-Cas2-Csn2 recognizes the leader-repeat sequence in the CRISPR array and integrates the acquired proto-spacer there (**Figure 3A**). Each integration event is followed by the creation of a new spacer-repeat unit, through the duplication of an existing repeat. In the second stage, a long primary transcript of the CRISPR *locus* (pre-crRNA) containing all the repeats and spacers is transcribed, along with the tracrRNA. Posteriorly, the tracrRNA binds to the pre-crRNA, acting as a guide for the recruited RNase III which cleaves the pre-crRNA in the presence of Cas9 into small mature crRNAs, containing a single repeat-spacer sequence. Upon completion of crRNA maturation, an endoribonuclease complex comprised of Cas9, tracrRNA and crRNA is formed (**Figure 3A**). This tracrRNA:crRNA-Cas9 complex locates and recognizes a foreign PAM in the viral sequence, binding to upstream DNA sequence through Watson and Crick base-pairing and triggering the degradation of the foreign DNA by Cas9-mediated DSB three base pairs upstream of the PAM

sequence (**Figure 3B**). Notably, since the PAM sequence is not included in the integrated spacers, this PAM-dependent recognition process constitutes a mechanism of prokaryotic "non-self-activation" of the adaptive immune system, preventing the induction of DBSs in their own CRISPR *locus* (Jinek et al 2012, Makarova et al 2011, Ran et al 2015, Rath et al 2015).

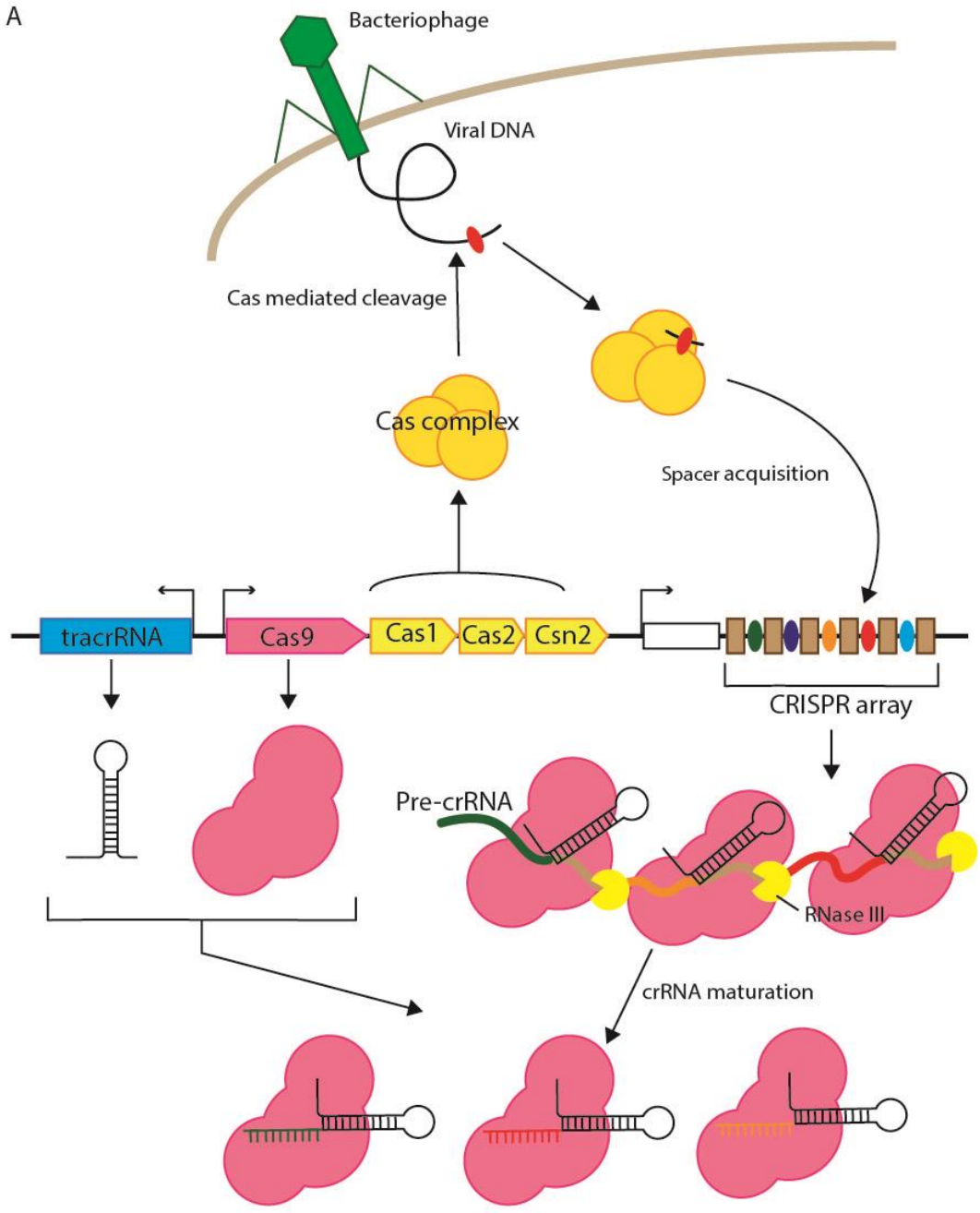

Figure 3- Type II CRISPR/Cas mediated prokaryotic immunity (figure continues below)

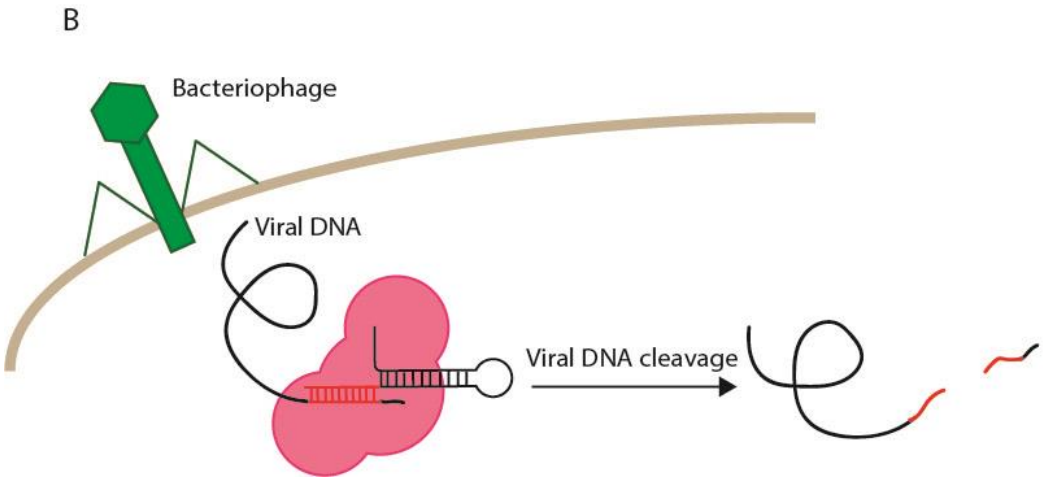

Figure 3- Type II CRISPR/Cas mediated prokaryotic immunity. A) During adaptation, the first stage of the immune response against bacteriophages, a Cas protein complex (Cas1-Cas2-Csn2) recognizes and cleaves viral DNA injected into the cell, and a resulting fragment named proto-spacer is integrated into the CRISPR array. In the expression stage, the CRISPR array is transcribed into pre-crRNA and undergoes maturation through the action of tracrRNA and Cas9, which promote RNAse III-induced cleavage of the pre-crRNA into small mature crRNAs. Following maturation, a mature crRNA forms a complex with Cas9 and tracrRNA (tracrRNA:crRNA-Cas9). **B)** Finally, in the interference stage, the tracrRNA:crRNA-Cas9 complex locates and recognizes a PAM element downstream a sequence present in an invading viral DNA that is complementary to the crRNA. The crRNA binds that sequence through Watson and Crick base-pairing, and its degradation is enacted by Cas9-mediated DSB.

In a gene editing context, Cas9 endonucleases can be programmed to cleave a specific DNA site using a single RNA duplex formed by the tracrRNA and the crRNA responsible for targeting and binding to the target 20 nucleotide DNA sequence. Even though the tracrRNA does not bind to the target DNA site, it was proven to be essential for crRNA maturation, which must occur in order for Cas9 DNA cleavage to be triggered. Furthermore, both the GG dinucleotide-containing PAM sequence adjacent to the crRNA and a seed sequence within the crRNA are required for an efficient recognition and cleavage of the target double-stranded DNA by Cas9 (Jinek et al 2012). However, since NGG is a very short and fairly common motif in genomes, this endonuclease platform can be used to target virtually any genetic region. This permissive PAM requirement constituted yet another advantage that contributed to the diffusion of gene editing strategies based on the type II CRISPR system of *S. pyogenes* (Adli 2018).

Importantly, the secondary structure of the tracrRNA:crRNA duplex suggested that the features required for DNA cleavage to occur, could be conserved in a single chimeric RNA molecule. Taking this into consideration, the tracrRNA:crRNA duplex was manipulated in order to generate a single RNA molecule capable of forming a complex with Cas9. This engineered chimeric RNA molecule consists in the crRNA 3´ end fused to the tracrRNA 5´,

forming a hairpin structure that retains the base-pair interactions naturally occurring between crRNA and tracrRNA. The development of this single-strand RNA molecule, named single-guide RNA (sgRNA), was a tremendous advance for programmed DNA cleavage and genome editing, since it further simplified the system, from three components to two (Jinek et al 2012).

1.3.4.2. Editing genes with CRISPR/Cas

CRISPR/Cas systems, unlike the other three nuclease platforms previously discussed, do not require protein engineering to target each novel DNA site. Due to their fixed DNA binding sequence requirements, size and repetitive composition, previous nucleases require a time consuming and expensive process to engineer large-scale protein libraries. The development of sgRNAs enables CRISPR/Cas systems to target new sites with an unprecedented ease by simply altering 20 nucleotide region within the sgRNA, which dictate binding specificity. Furthermore, the Cas9 protein and the sgRNA are not directly coupled (**Figure 4A**), making for a system capable of multiplexing by using multiple sgRNAs to target several *loci* and guide multiple DSBs. These facts makes the CRISPR/Cas system a highly attractive and scalable platform, that any research laboratory can easily get into, in order to develop strategies for inactivating a gene *locus*, through the introduction of indels, or mediating substitutions, inversions, duplications, knock-ins, and other complex modifications in the genetic code (Fernandez et al 2017, Maeder & Gersbach 2016).

Le Cong and co-workers from Feng Zhang's group, who had worked with other gene editing systems such as TALENs, were pioneers in successfully harnessing this RNA-programmable nuclease system to perform gene editing in eukaryotic cells, namely in human and mouse cells. In order to ensure that the Cas9 protein would be translocated to the nucleus, two nuclear localization signals were attached to the protein, providing an increased efficiency when targeting a particular genetic *locus* (the human *EMX1 locus*). Initially, human embryonic kidney (HEK) 293T and mouse Neuro-2a cells were both co-transfected with tracrRNA, pre-crRNA and *S. pyogenes* Cas9 (SpCas9), demonstrating that targeted genome DSBs were achievable in different *loci* across multiple organisms with efficiencies comparable to those of TALENs, which at the time were the golden standard gene editing platform. In addition, in the same study Feng Zhang's group was able to introduce restriction sites for *Hind*III and *Nhe*I into the *EMX1 locus* and perform targeted deletion in two target sequences within the same *locus*, through multiplexing of the CRISPR/Cas system (Cong et al 2013).

Concurrently, a different research group used a similar approach to that of Feng Zhang's group, with the intent of testing the editing capabilities of the CRISPR/Cas9 system. They used engineered HEK 293T cells containing a genomically-integrated green fluorescent protein (GFP) gene disrupted by a STOP codon and a genomic fragment from the AAVS1 *locus*, which prevented fluorescence from the translated GFP protein. Upon transfection of the reporter cells with the CRISPR/Cas system containing a sgRNAs targeting the interfering AAVS1 fragment and a repair template capable of reinstituting the normal GFP sequence, rapid restoration of GFP fluorescence was observed in comparison to the TALEN control, reinforcing that RNA-guided Cas9 nucleases were able to efficiently promote HDR in target sites. Similarly to what Le Cong and collaborators described, this research group was also able to perform high-efficiency deletions via NHEJ when co-transfecting different human cell lines with guide RNAs targeting different *loci*, demonstrating the feasibility of CRISPR/Cas9 multiplexing. Finally, the research team integrated a DNA sequence containing a puromycin resistance gene and another containing the GFP gene at the AAVS1 *locus* from HEK 293T cells and iPSCs, proving the ability of the CRISPR/Cas9 system to integrate foreign DNA at an endogenous *locus* in human cells (Mali et al 2013b).

The years following the first description of gene editing using the Cas9-sgRNA system saw a "boom" in the interest in this tool, resulting in hundreds of papers using this system being published within a short time span (Adli 2018). The versatility and accessibility of the Cas9 molecular scissor took gene editing further into the central stage of Life Sciences. With the growing interest and usage of the CRISPR/Cas system in Molecular Biology, the scientific community developed new ways to optimize the strategies performed in the two revolutionary papers described above. A new wave of strategies that take advantage of the features of Cas9 and guide RNA molecules has gained *momentum*, creating CRISPR/Cas9-based tools which perform gene editing without relying on DSB or that perform functions other than gene editing, such as gene expression regulation.

1.3.4.3. The potential of CRISPR/Cas9 beyond DNA double-strand breaks

Even prior to the first publications using CRISPR/Cas9 as a tool for gene editing, this system was modified and adapted. The first adaptations came during the description of the complex inner workings of the Cas9-sgRNA ribonucleoprotein, namely when determining which domains were responsible for the interference process, by mutating active-site residues in different Cas9 endonuclease domains. Cas9 proteins contain domains that are homologous to HNH and RuvC endonucleases; due to Cas9 folding, these domains mediate the cleavage of opposing DNA strands, generating a DSB. The RuvC-like nuclease domain, localized near the amino terminus, cleaves the complementary (+) DNA strand, whereas the HNH-like nuclease domain, localized in the middle of the protein, cleaves the noncomplementary (-) DNA strand. Taking this into account, researchers replaced active-site residues in the RuvC or in the HNH-like motifs for alanine (A) amino acids via site-directed mutagenesis. Mutation of a single amino acid was able to transform the Cas9/sgRNA complex into a strand-specific nickase that cleaves a single DNA strand. In the case of RuvC mutants (D31A or D10A), the DNA would be cleaved solely in the complementary strand, in contrast to HNH mutants (N891A or H840A), that cleaved only the noncomplementary strand (Gasiunas et al 2012, Jinek et al 2012).

Abolishing the catalytic activity of both Cas9 endonuclease domains through the same strategy gave rise to a catalytically inactive Cas9 (dead Cas9; dCas9). Mutations in RuvC and HNH have no significant effect on the Cas9/sgRNA complex DNA-binding properties since these mutant proteins retain the same affinity for DNA as the wild-type protein. Therefore, inactivating both endonuclease sites leads to a Cas9 protein completely absent of catalytic activity, but with conserved DNA binding properties driven by the sgRNA (Gasiunas et al 2012, Jinek et al 2012).

Taking advantage of dCas9 DNA targeting and binding abilities, reminiscent of what was previously done with ZFNs and TALENs, dCas9 was fused with transcriptional activators or repressors. The resulting fusion proteins allow a fine regulation of gene expression, without the need to modify the genomic content (**Figure 4B**).

Following an initial study with modest results concerning the degree of transcriptional repression, Gilbert and colleagues sought to improve the efficiency of CRISPR interference (CRISPRi), that is, using the CRISPR technology to repress gene expression in a sequence-specific manner, by fusing dCas9 with a transcriptional repressor (not to be confused with the interference step of the native CRISPR adaptive immune system). The dCas9 protein was fused

to the Krüppel-associated box (KRAB) domain, which is able to promote gene silencing through the recruitment of chromatin-modifying machinery, to form a chimeric protein capable of inducing targeted gene repression. To address the efficiency of CRISPRi silencing towards endogenous human genes, sgRNAs targeting two cell surface transmembrane protein-coding genes - the transferrin receptor (CD71) and C-X-C chemokine receptor type 4 (CXCR4) - were designed and transfected into HeLa cells stably expressing dCas9-KRAB. Stable expression of dCas9-KRAB and simultaneous transfection of both sgRNAs (multiplexing) was also performed. In these experiments, the endogenous expression of CD71 and CXCR4 was efficiently knocked down in all the different sgRNA expression scenarios, either by single transfection or co-transfection (Gilbert et al 2013).

Similarly, dCas9 can also be fused to transcriptional activators. Indeed, Chavez and collaborators were able to fuse a VPR activator complex (Vp64, p65 and Rta) to dCas9. Using sgRNAs designed to target a set of genes related to cellular reprogramming, development, and gene therapy, the dCas9-VPR complex proved to be able to induce robust expression of its targets within their native chromosomal context, in HEK 293T cells (Chavez et al 2015).

Apart from fusing dCas9 with transcriptional activators or repressors, scientists have also sought to alter gene expression profiles by modulating the chromatin condensation state, by fusing enzymes responsible for direct epigenetic modifications (**Figure 4C**). In mammalian cells, methylation of CpG-rich promoter regions correlates with their transcriptional repression. With that in mind, DNA methyltransferases Dnmt3a and Dnmt3L were engineered into a single chain (Dnmt3a3L) and fused to dCas9. Targeting the fusion to a promoter of interest led to an increase in CpG methylation patterns, thus demonstrating epigenome edition (Stepper et al 2017). Additionally, it has been demonstrated that epigenome modifications can also be performed at the histone level using the same rationale. Such modifications consist in the removal of methyl or the addition of acetyl groups to the histone tails, which can be perform by a dCas9 protein fused to a histone demethylase LSD1 or a histone acetyltransferase p300, respectively (Hilton et al 2015, Kearns et al 2015).

1.3.4.4. The endless potential of CRISPR/Cas9 fusion proteins

Base editing is an alternative to "traditional" genome editing, whereby one base pair HDR directed repair, DSBs or donor DNA templates. At the same time, the process minimizes stochastic editing outcomes such as undesired indels with added efficiency. Targeted base editing is made possible by fusing Cas9 nickase (nCas9) with activation-induced cytidine deaminase (AID) or tRNA adenine deaminase (TadA) (**Figure 4D**). Upon nicking of the DNA by the nCas9 protein, the deaminases exchange a cytosine for a thymine (C:G to T:A) or an adenine for a guanine (A:T to G:C), respectively, in the upstream vicinity of the PAM (Gaudelli et al 2017, Nishida et al 2016).

Several strategies for deliberately restraining Cas9 activity have been developed, creating inducible systems that allow for a temporal regulation of Cas9 activity and thereby decrease off-target genome modifications. Several of these strategies rely on ligand-biding-regulated systems (**Figure 4E**). In one of these systems, Cas9 activity is regulated by manipulating its degradation rate through a Cas9 chimeric construct containing an unstable dihydrofolate reductase (DHFR) domain, that marks the chimeric complex for degradation. Binding of trimethoprim (TMP) to the DHFR domain stabilizes it, halting the complex degradation and allowing for Cas9 to remain active (Maji et al 2017). Also taking advantage of this type of chemical compound-binding process, the activity of Cas9 can be restrained by the absence of 4-hydroxytamoxifen (4-HT). To achieve this, a ligand-dependent intein was inserted into Cas9, rending it inactive. 4-HT binding to the intein triggers protein splicing, which restores Cas9 activity (Davis et al 2015). Combining a split fragment strategy with ligand binding, two split Cas9 fragments were engineered, one with a FK506 binding protein 12 (FKBP) in the C-terminal fragment (Cas9(C)-FKBP) and another with a FKBP rapamycin binding (FRB) domain in the N-terminal fragment (Cas9(N)-FRB) (**Figure 4F**). Upon addition of rapamycin, (Cas9(C)-FKBP) and (Cas9(N)-FRB) dimerize, causing the complex to become active and to be translocated into the nucleus (Zetsche et al 2015).

Besides the temporal control over Cas9, optical regulation of light-activated proteins function enables precise spatial control over gene editing. Optochemical control over the CRISPR/Cas9 system is achieved through the incorporation of a photocaged lysine (PCK) at position 866 of Cas9. PCK is inactive or "caged" until the protein is exposed to UV radiation, at which point PCK is converted into a lysine residue, generating Cas9 capable of performing DSBs (Hemphill et al 2015). Notably, the combination of optically controlled Cas9 and a split

fragments strategy gave rise to two split Cas9 fragments containing photoswitchable proteins from the Magnet system (**Figure 4G**). The N-terminal fragment was fused to a positive magnet (pMag) and the opposing C-terminal fragment was fused to a negative magnet (nMag). Upon blue light stimulation, both magnets fuse, forming a reversible heterodimer and inducing Cas9 activity (Nihongaki et al 2015).

Most recently, a Cas9-based strategy that promotes genome editing without DBS or donor DNA, much like what has been achieved with base editing, was engineered. This strategy, termed Prime editing (**Figure 4H**), directly introduces new genetic material into a target site, using a nCas9 protein fused to an engineered M-MLV reverse transcriptase (Prime editor). The fusion protein is complexed with an engineered chimeric sgRNA (prime editing gRNA) that targets Cas9 to the target DNA site, and that at the same time encodes for the desired edit, which is inserted through the activity of the reverse transcriptase (Anzalone et al 2019).

A Cas9-based nuclease system requiring simultaneous binding events from two distinct Cas9 monomers has also been engineered. Similarly to what was done in the past with dimeric ZFNs and TALENs, dCas9 was fused to Fok1 (**Figure 4I**). Despite having a lower efficiency compared to Cas9, this system displays increased specificity, since two 20 nucleotide sequences must be recognized for FokI to dimerize and become catalytically competent (Guilinger et al 2014).

Due to the high flexibility of the sgRNA molecule, Cas9 fusion proteins can be used in applications that go beyond gene editing and expression regulation. A CRISPR-based technique capable of labeling specific genomic elements *in vivo* was developed by fusing a fluorescent protein, such as green fluorescent protein (GFP) to dCas9 (**Figure 4J**). This live imaging technique provides a powerful tool to visualize the position of specific genes within the nucleus, as well as studying chromatin architecture and nuclear organization (Chen et al 2013).

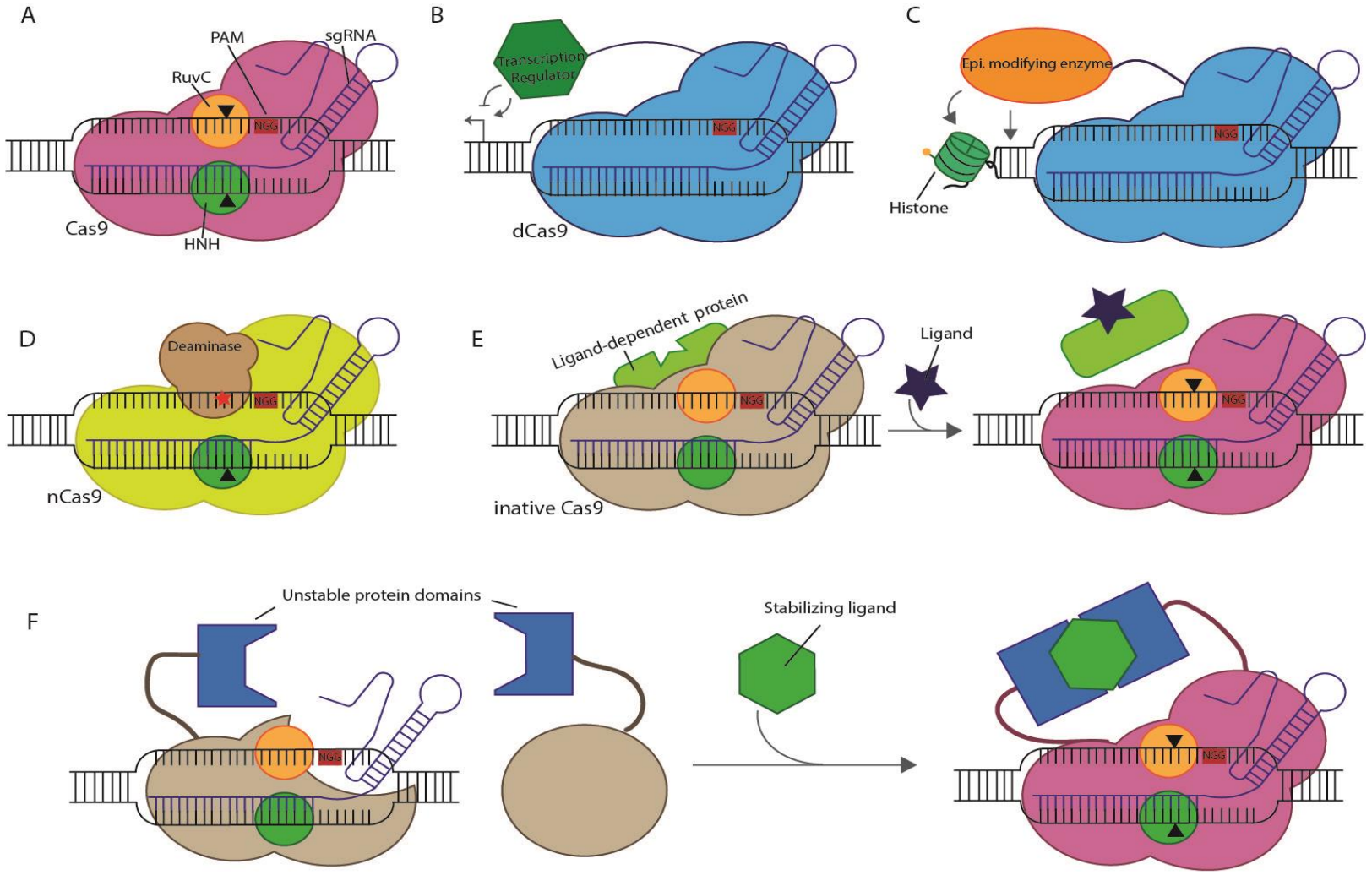

Figure 4- Different CRISPR/Cas9-based tools. (Figure continues below)

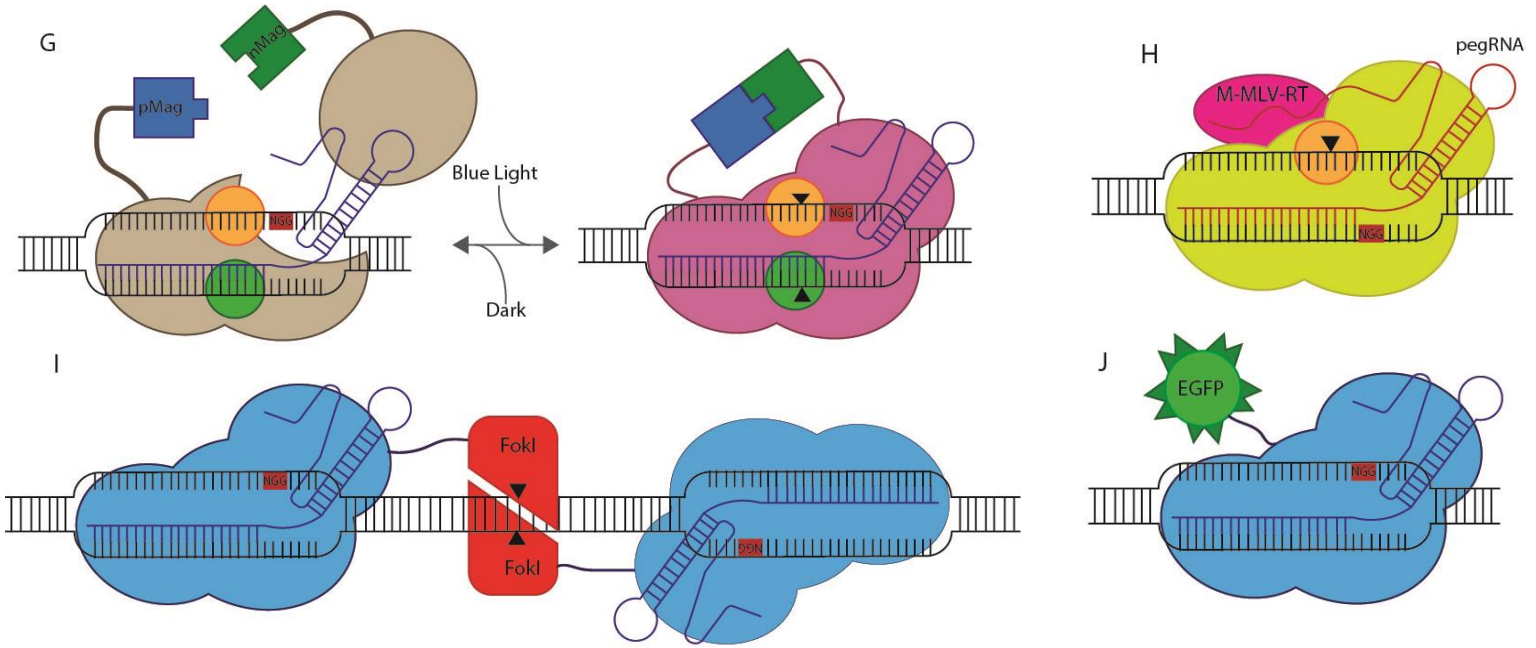

Figure 4- Different CRISPR/Cas9-based tools. A) The Cas9 ribonucleoprotein complex is constituted by the Cas9 protein (purple) and the sgRNA molecule (blue). Cas9 has two catalytic domains, RuvC (orange) and HNH (green), which induce DNA double stranded breaks (▼) 3 base pairs upstream of the PAM (5'-NGG-3', in the case of *S. pyogenes* Cas9) (red box). **B)** dCas9 (light blue) has both catalytic domains mutated, impairing the ability to induce DSBs. When fused to a transcription repressor (for example, KRAB) or a transcription activator (for example, the VPR complex), dCas9 can regulate target gene expression. **C)** dCas9 fused to epigenetic regulators such as DNA methylases (Dnmt3a3L) or histone modifiers (demethylase LSD1 and acetyltransferase p300) is able to alter the epigenome. **D)** Mutating one of the catalytic domains of Cas9 Cas9 transforms it into a nCas9 nickase. nCas9 fused to deaminases (AID or tadAD) is capable of editing bases without DSBs. **E)** Inducible inactive Cas9 (grey) fused to ligand-depend proteins such as DHFR or intein can be "switched on", upon administration of TMP or 4-HT, respectively. **F)** A split and inactive Cas9, with each terminal fragment fused to FKBP or FRB, dimerizes and becomes catalytically active after addition of rapamycin to the medium. **G)** Two inactive fragments of Cas9, each fused to photoswitchable proteins (pMag or nMag), dimerize in the presence of blue light, acquiring catalytic activity, and become inactive in the dark. This constitutes a real time-inducible Cas9 form. **H)** Fusing dCas9 with the restriction enzyme FokI, an enzyme that is catalytically activi only when forming a homodimer, enables the induction of DSB with increased specificity. **I)** nCas9 fused to an engineered M-MLV reverse transcriptase, complexed with an engineered sgRNA (prime editing gRNA) that targets a particular DNA *locus* and encodes for the desired edit, enables genetic insertions without the need for DSBs. **J)** Fusing dCas9 with EGFP enables precise live tracking of target DNA regions.

1.3.4.5. The CRISPR/Cas9 system as a tool to treat polyQ disorders

Alternatively to mRNA-directed gene silencing, gene editing tools such as the CRISPR/Cas9 system offer the possibility of countering the polyQ disease at the DNA level, the most upstream point of the pathological cascade. This may be accomplished through gene-specific knockouts, pre-transcriptional gene silencing or replacement/repair of mutations such as the CAG-expanded tract (Hsu et al 2014). Additionally, gene editing has been shown to efficiently induce modifications in a diverse set of cell types, in which neurons are included (Nishiyama et al 2017).

A proof-of-principle study consisting in an allele-specific inactivation of the *HTT* gene was performed using the CRISPR/Cas9 system. Following a screen of the *HTT locus* to identify *single-nucleotide polymorphisms* (SNPs) unique to the mutant allele, gRNAs targeting the identified SNPs were designed. Transfection of primary HD fibroblasts with two sgRNAs targeting a sequence in the upstream promotor region and intron 3 resulted in the excision of the promotor region, transcription start site and expanded CAG repeat. This led to a complete inactivation of the mutant allele in primary fibroblasts from an HD patient, while the normal allele remained active (Shin et al 2016).

CRISPR/Cas9-mediated deletion of the expanded CAG repeat of the *ATXN3* gene was also performed in induced pluripotent stem cells (iPSCs) derived from SCA3 patients. In this study, SCA3-iPSCs were transfected with two sgRNAs targeting sequences flanking the polyQ-encoding region. Results showed that the exon 10 containing the CAG repeats was deleted, while exon 9 and exon 11 remained unchanged. Furthermore, following gene editing, the ubiquitin-binding capacity of ataxin-3 was maintained. Importantly, corrected SCA3-iPSCs were able to differentiate to both neural stem cells and neuronal cells, implying pluripotency was retained after treatment (Ouyang et al 2018).

These and other observations suggest that the unprecedented possibilities offered by the CRISPR/Cas9 gene editing systems may be directed to the development of innovative gene therapy strategies for these and other polyQ diseases, including SCA2.

CHAPTER 2 - OBJECTIVES

Currently, there is no therapy able to delay or stop SCA2 progression. As a polyQ disorder with a monogenic origin, SCA2 represents a well-suited candidate for gene therapy-based approaches, namely those aiming at directly silencing the causative mutant gene. Until now, silencing approaches that have been investigated do not inhibit the transcription of the CAG-expanded mRNA; instead, they rely on targeting *ATXN2* mRNA itself, which does not eliminate possible RNA-derived toxicity. For example, a therapeutic strategy targeting *ATXN2* mRNA using antisense oligonucleotides, that partially suppress ataxin-2 translation, has been developed in SCA2 mice models (Scoles et al 2017). However, evidence has repeatedly shown that mRNA derived from expanded polyQ protein genes exhibits neurotoxic activity (Wang et al 2011).

In this project, we aimed to design and develop a CRISPR/Cas-based strategy for SCA2 therapy, based on *ATXN2* silencing. To achieve this goal we tested two different approaches: we employed CRISPRi to repress endogenous wild-type *ATXN2* expression at a pre-transcriptional level, and alternatively DSB-competent CRISPR/Cas9 to permanently impair translation of ataxin-2. The first approached entailed the development of a strategy targeting the *ATXN2* promoter and which impaired the initiation of gene transcription. This would repress gene expression directly at the genomic level, before toxic expanded RNA is formed and thus countering the disease at the most upstream point of the pathogenic cascade. The second strategy would be capable of introducing a premature stop codon upstream of the CAG tract located in exon 1 of *ATXN2*, therefore impairing the translation of the protein at an earlier step of its synthesis.

CRISPR/Cas-mediated gene knock-out relies on the NHEJ DNA repair system to introduce an indel mutation that leads to the formation of a premature stop codon. In order to induce a NHEJ-stimulating DSB upstream of the CAG tract, one sgRNA was designed to target a sequence in between the translation initiation codon and the beginning of the CAG repeat region. The sgRNA was cloned into an expression plasmid containing catalytically active Cas9, and the construct was transfected into HEK 293T cells. In order to assess if indels were generated at the intended site through this approach, genomic DNA was extracted and the *ATXN2* gene was sequenced. The effects of the on *ATXN2* expression were evaluated at RNA and protein levels, by qPCR and Western blot, respectively.

Aiming to silence *ATXN2* expression through CRISPRi, several sgRNAs that would be able to direct a protein fusion, comprised of dCas9 and a KRAB transcription repressor, to

promoter sequence of *ATXN2* were designed. Binding of dCas9-KRAB to *ATXN2* promoter would enable *ATXN2* silencing through the KRAB-mediated recruitment of chromatin-modifying machinery and a resulting condensation of the chromatin. Upon cloning of the sgRNAs into an expression plasmid codifying the dCas9-KRAB fusion, HEK 293T cells were transfected and the effect of this approach in repressing *ATXN2* expression, both at the RNA and protein levels, were evaluated, by qPCR and Western blot, respectively.

CHAPTER 3 - MATERIAL AND METHODS

3.1. Plasmids

lentiCRISPR v2

lentiCRISPR v2 (Addgene plasmid # 52961), engineered by (Sanjana et al 2014) was a gift from Dr. Feng Zhang. It encodes for a humanized *S. pyogenes* Cas9; a sgRNA scaffold-region in which the customizable sgRNA-codifying oligonucleotides can be inserted and transcribed under the control of the hU6 promotor; and a puromycin resistance cassette, all inserted into a lentiviral backbone (**Figure 5**).

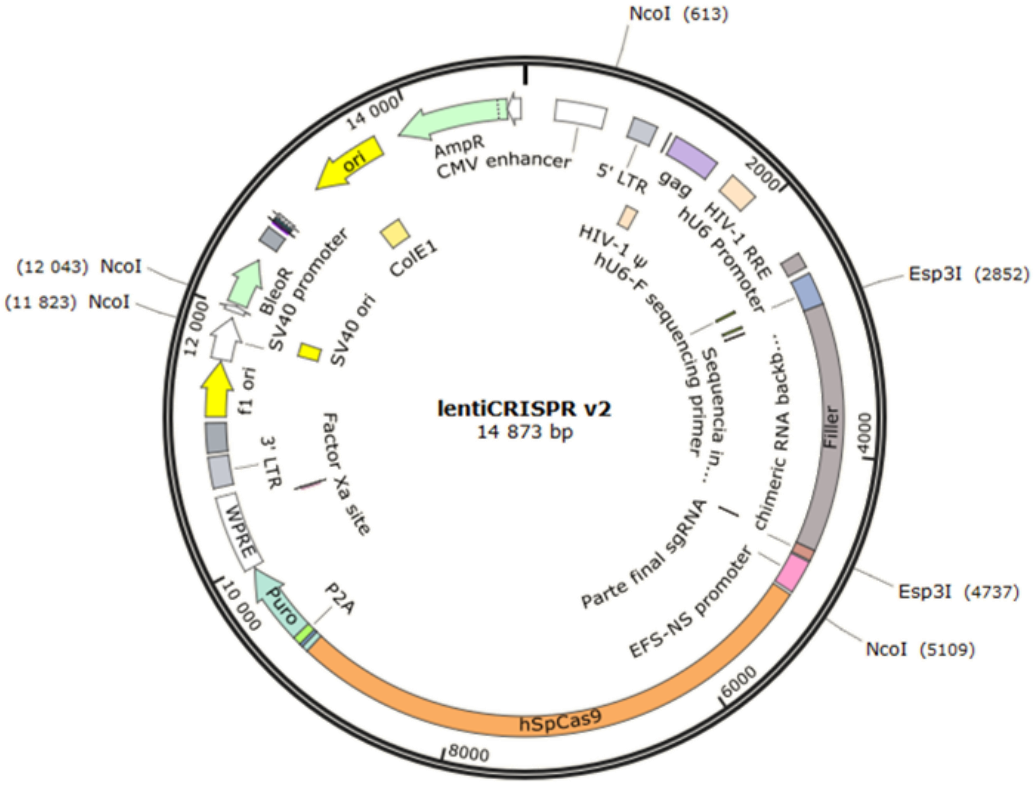

Figure 5- Map of the lentiCRISPR v2 plasmid. The plasmid has an origin of replication with an EFS-NS promoter for the humanized *S. pyogenes* (SpCas9). The sgRNA insert is expressed under the control of a hU6 promotor. Resistance cassettes for ampicillin and puromycin are also present in this plasmid. The size of the plasmid in base pairs, as well as the plasmid´s name, are shown at the center (image generated using SnapGene 1.1.3).

pLV hU6-sgRNA hUbC-dCas9-KRAB-T2a-Puro

pLV hU6-sgRNA hUbC-dCas9-KRAB-T2a-Puro (Addgene plasmid # 71236), engineered by (Thakore et al 2015), was a gift from Dr. Charles Gersbach. This plasmid codifies for a dCas9-KRAB fusion protein, comprised of an inactive humanized *S. pyogenes* Cas9 protein (dCas9), fused in frame with KRAB at the C-terminal; a sgRNA scaffold-region in which the customizable sgRNA-codifying oligonucleotides can be inserted and transcribed under the control of the U6 promotor; and a puromycin resistance cassette, all inserted into a lentiviral backbone (**Figure 6**).

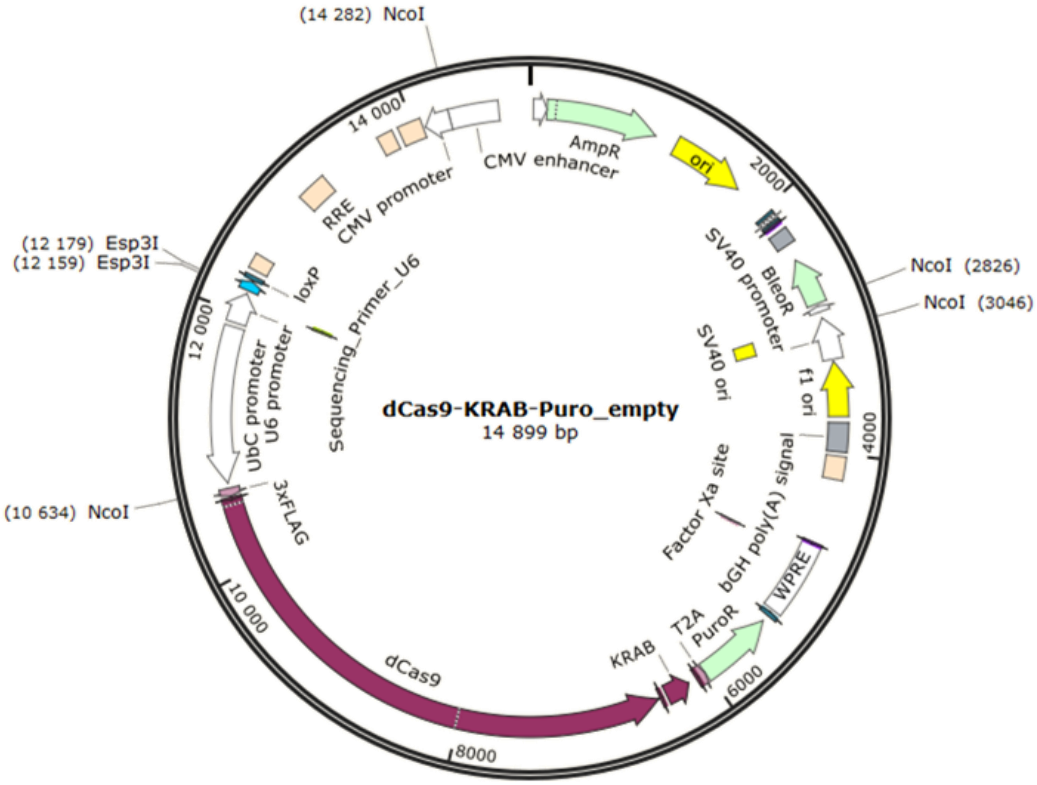

Figure 6- Map of the pLV hU6-sgRNA hUbC-dCas9-KRAB-T2a-Puro. The plasmid has an origin of replication with an UbC promoter for the dCas9-KRAB. The sgRNA insert is expressed under the control of a U6 promotor. Resistance cassettes for ampicillin and puromycin are also present in this plasmid. The size of the plasmid in base pairs, as well as the plasmid´s name, are shown at the center (image generated using SnapGene 1.1.3)

3.2. Plasmid amplification

Competent 5-alpha *E. coli* (New England Biolabs, USA) were transformed with the lentiCRISPR v2 or dCas9-KRAB plasmids through heat shock. Cells were grown at 37 ºC in SOC outgrowth medium (2% vegetable peptone, 0,5% yeast extract, 10 mM NaCl, 2,5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$ and 20 mM glucose) for 1 h, then plated in dishes containing LB-agar with ampicillin. Cells were incubated overnight at 37 ºC and the resulting colonies were inoculated into lysogeny broth (LB; NZYTech, Portugal) and grown overnight with orbital rotation at 37 ºC. Following incubation, plasmidic DNA was purified using a NucleoBond® Xtra Midi Plus EF kit (Machery-Nagel, Germany), according to the manufacturer's recommendations. Upon purification, DNA concentration and purity were assessed using Nanodrop 2000c (Thermo Scientific, USA).

3.3. SgRNA design

The DNA sequence of the human *ATXN2* gene (NG_011572.3) was imported from the National Center for Biotechnology databank (NCBI, USA; https://www.ncbi.nlm.nih.gov/). Annotation for the *ATXN2* promoter was based on the Eukaryotic Promoter Database (EPD; Swiss Institute for Bioinformatics; https://epd.epfl.ch/human/human_database.php). *ATXN2* has two in-frame putative start codons, and the one located four codons upstream of the CAG tract was the one selected as reference for the design, based from the cross analysis of information from NCBI, EPD and other bibliographical sources (Scoles & Pulst 2018).

The bioinformatic platform Benchling (https://www.benchling.com/) was used to scan the *ATXN2* sequence and detect SpCas9 PAMs (NGG) and the corresponding putative sgRNA sequences in a region spanning from the middle of the promoter region to immediately upstream of the CAG tract. From among the candidate sgRNA sequences, the most promising ones were chosen based on their target sites and the On and Off target scores attributed by the platform, based on the model described in (Doench et al 2016) (a high Off-target score, corresponding to a low number of predicted off-targets, was prioritized). sgRNAs carried by dCas9-KRAB complexes were selected from among those targeting the promoter region, while a sgRNA to be carried by Cas9 was selected considering that it targeted a region between the translation

start site and the CAG tract (**Figure 7**). A sgRNA sequence targeting *Escherichia coli* LacZ that has been validated as having no targets in human cells was used as a cut-inefficient control (Platt et al 2014).

3.4. SgRNA cloning

The 20-nucleotide sequences of the sgRNAs, responsible for target binding, were cloned in the form of double-stranded DNA oligonucleotides into their respective expression vector - dCas9-KRAB or lentiCRISPR v2. Each oligonucleotide strand (template and codifying) were complementary to each other, with the exception to an overhang region located at the 5' end of each strand, comprised of four nucleotides, which enabled the introduction of the oligonucleotide into a BSmBI (Esp3I) restriction site present in both plasmids, downstream of the hU6 promotor and upstream of the sgRNA scaffold-codifying region. For sgRNA sequences lacking a guanine nucleotide in the 5' end of the codifying strand, between the overhang region and the 20 nucleotides which target the human *ATXN2* gene, a guanine was added in order to enable the polymerase III-dependent U6 promoter to function. The sgRNA oligonucleotides were acquired from *Invitrogen* (Thermo Fisher, USA).

Phosphorylation and annealing of the oligonucleotide pairs was performed using T4 polynucleotide kinase (Thermo Scientific, USA) following the manufacturer's recommendations. Expression plasmids were digested individually with BmsBI (Esp3I; Thermo Scientific, USA) and purified upon 1% (w/v) agarose gel electrophoresis and band excision with the NZYGelpure kit (NZYTech, Portugal). The annealed sgRNA oligonucleotides were cloned into the open expression plasmids, through a ligation reaction using T7 ligase (New England Biolabs, USA), according to the manufacturer's recommendations.

Products resulting from these ligations were transformed into competent 5-alpha *E. coli* and grown in LB containing ampicillin, at 37 ˚C. Isolated cell colonies were inoculated into LB medium and the plasmidic DNA of each individual culture was purified with the NZYMiniprep kit (NZYTech, Portugal). Success of the cloning procedure was accessed through Sanger sequencing (U6 forward sequencing primer; GATC biotech, Germany).

The quality and integrity of the cloned plasmids was evaluated though enzymatic restriction with NcoI restriction enzyme (Thermo Scientific, USA). Resulting fragments were loaded into a 1 % (w/v) agarose gel and separated by electrophoresis, during 60 minutes at 90 V. The Gel was then scanned using a ChemiDoc™ XRS+ System (Bio-Rad, USA). SnapGene 1.1.3 was used to calculate the expected size of fragments resulting from plasmids enzymatic digestion.

3.5. Cell culture and transfection

Human embryonic kidney (HEK) 293T cells were cultured in Dulbecco´s Modified Eagle Medium High glucose (DMEM HG; BioConcept, Switzerland) with L-Glutamine, containing 3,7 g/L sodium bicarbonate (NaHCO3; Fisher Scientific, USA) and 10 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES; Fisher Scientific, USA) and supplemented with 10% (v/v) fetal bovine serum (FBS; Gibco, USA) and 1% (v/v) penicillin-streptomycin (Gibco, USA). Cells were grown in 75 cm² T-Flasks (SPL Life Sciences, South Korea) in an incubator at 37 ºC, with a humidified atmosphere containing 5% CO_2**.** Every 2 days, upon reaching an average confluence ≥ 80 %, cells were split, and the media renewed.

Transfection with polyethylenimine **(**PEI; Polysciences, USA) was carried out following the manufacturers specifications. In the day prior to the transfection, confluent cells were diluted 1:6 and seeded into Costar® 6 Well Multiwell Plates (Corning Incorporated, USA). Seeded cells were transfected with 500 µg of Plasmid DNA and 3 µL of PEI per well. After 24 h, the transfection medium was substituted by DMEM with 10 µg/mL of puromycin and the next by fresh culture medium. Following a 24 h incubation with fresh DMEM, cells were collected for DNA or RNA purification, or for cell lysate preparation.

3.6. RNA extraction

Cultured cells were washed with cold phosphate buffer saline (PBS; 138 mM NaCl, 2,7 mM KCl, 11,76 phosphate buffer, pH 7,4) and immediately frozen at -80 ºC. Posteriorly, cells were disrupted with β-mercaptoethanol and vigorously scraped while on ice. Total RNA was extracted using the NZY Total RNA Isolation kit (NZYTech), according to the manufacturer's

recommendations. Upon assessing RNA concentration and purity using Nanodrop 2000c (Thermo Scientific, USA), samples were immediately frozen at -80 ºC.

3.7. Preparation of cell lysates

Cultured cells were washed with cold phosphate buffer saline (PBS; 138 mM NaCl, 2,7 mM KCl, 11,76 phosphate buffer, pH 7,4) and immediately frozen at -80 ºC. Posteriorly, cell lysates were prepared by adding radioimmunoprecipitation buffer (RIPA; Milipore, USA), supplemented with a mini protease inhibitors mixture tablet (Roche, Switzerland). 100 µL of supplement buffer was added per well and cells were scraped vigorously while on ice. Then cell lysates were sonicated on a Bioruptor® Pico sonication device (Diagenode, Belgium) for 5 cycles, each with 30 seconds ON and OFF. Sonicated samples were centrifuged at 14.000 rpm for 15 mins, at 4 ºC, and the supernatant recovered. Protein concentration was determined by the bicinchoninic acid (BCA) method using the Pierce® BCA Protein assay kit, according to the manufacturer's recommendations (Thermo Scientific, USA). Lysates were kept frozen at -20 ºC.

3.8. Western blot

Cell lysates were diluted in RIPA buffer according to their protein concentration, so that all protein samples separated through electrophoresis would have a total amount of 50 µg of protein and the same final volume. Laemmli SDS sample buffer, reducing (4X; 250 mM Tris-HCl (pH 6.8), 8% (w/v) SDS, 40% (w/v) glycerol, 8% (w/v) β-mercaptoethanol, and 0.02% (w/v) bromophenol blue) was added and samples were heated for 5 mins at 95 ºC. Protein samples were loaded into 4% (w/v) bis-acrylamide stacking gels (0.5 M Tris-HCl, pH 6.8 with 0.4% (w/v) SDS) and separated in 8 % (w/v) resolving gels (1.5 M Tris-HCl pH 8.8 with 0.4% (w/v) SDS) through electrophoresis, until the NZYcolour Protein Marker II (NZYTech) reached the desired band separation. Electrophoresis was performed at 80 V until the samples reached the running gel, and afterwards at 120 V. Proteins were electrotransfered at 500 mA into polyvinylidene fluoride (PVDF) membranes for 3 h 30 mins, by wet transfer at 4 ºC.

Membranes were blocked with Tris-buffered saline (20 mM Tris and 150 mM NaCl, pH 7,6) containing 0,1% (v/v) Tween 20 (TBS-T) and 3% (w/v) BSA, for 1 hour with agitation. Primary antibodies were diluted in TBS-T with 3% (w/v) BSA, and then membranes were incubated in the appropriate primary antibody (**Table I**) solution, overnight at 4 ºC. Membranes were washed 3 times with TBS-T, during a period of 10 minutes for each wash. Afterwards, membranes were incubated with the suitable secondary antibody, also diluted in TBS-T with 3% (w/v) BSA. Following 2 h of incubation with the secondary antibody (**Table I**), the washing process was repeated (3 times, 10 minutes each) using TBS-T. Finally, membranes were resolved using Amersham ECL Prime Western Blotting Detection Reagent (GE Healthcare, USA) and scanned with a ChemiDoc™ XRS+ System (Bio-Rad, USA).

Antibody reprobing was performed by washing the membrane for 3 h in TBS-T, followed by the labelling process previously described. Densitometric analysis of the scanned Western blots was performed using ImageJ v1.52 software (National Institute of Health, USA).

Western blot antibodies

Antibodies used in Western blot procedures are presented on Table I.

Table I: Antibodies used in Western blot analysis.

	Antibody name	Species	Dilution	Incubation	Manufacturer
Primary antibodies	Purified Mouse Anti-Ataxin 2	Mouse monoclonal	1:1000	Overnight, 4 ºC	BD transduction laboratories (USA)
	Anti-GAPDH (D16H11)	Rabbit monoclonal	1:1000	Overnight, 4 ºC	Cell signaling technology (USA)
Secondary antibodies	ECL Anti-mouse IgG	Mouse polyclonal	1:10.000	2 h, RT	GE Healthcare (USA)
	ECL Anti-rabbit IgG	Rabbit polyclonal	1:10.000	2 h, RT	GE Healthcare (USA)

Abbreviations: RT, room temperature

3.9. qPCR

cDNA was obtained by reverse transcription of the purified RNA samples using the iScript cDNA synthesis kit (Bio-Rad, USA), following the recommended protocol. qPCR was performed using SsoAdvanced Universal SYBR Green Supermix kit (Bio-Rad, Portugal) according to the manufacturer's recommendations. cDNA amplification was carried out using a pair of primers (Qiagen, USA) that targeted endogenous ATXN2 in HEK 293T cells (**Table II**). As control, a pair of primers targeting the human GAPDH gene was used (Merck, Germany;). qPCR was performed in a CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad, USA).

Table II: Primers used in qPCR procedures.

Primer name	Sequence	Manufacturer
FH1_GAPDH	5'-ACAGTTGCCATGTAGACC-3'	Merck
RH1_GAPDH	5'-TTGAGCACAGGGTACTTTA-3'	Merck
ATXN2_2.2	Not disclosed	Qiagen

3.10. *ATXN2* gene sequencing

Genomic DNA from HEK 293T cell cultures was extracted and purified with the GeneJET Genomic DNA Purification Kit (Thermo Scientific, USA), following the manufacture's recommendations. DNA concentration and purity were evaluated using Nanodrop 2000 (Thermo Scientific, USA).

DNA samples were submitted to PCR using Phusion™ High-Fidelity DNA Polymerase (Thermo Scientific, USA) and a pair of custom primers (Table 3) that amplified a region of the human *ATXN2* with approximately 950 bp, comprising the Cas9/Indel target site. PCR parameters were set according to the DNA polymerase manufacturer and the primers melting temperature. Resulting PCR amplicons were loaded into a 1,5 % (w/v) agarose gel and separated through electrophoresis, during 50 minutes at 90 V. Gels were then scanned using a ChemiDoc™ XRS+ System (Bio-Rad, USA) and the bands of interest were excised from the agarose gel using a sterile scalpel. The DNA in the excised gel fragment was purified by the NZYGelpure kit (NZYTech, Portugal), following the recommended protocol, and its sequence accessed through Sanger sequencing (forward primer; GATC biotech, Germany).

3.11. Statistical analysis

Statistical analysis was performed using the GraphPad 7.04 software (GraphPad Software, USA). Comparison between two groups was carried out using unpaired Student's t-test. Multiple comparisons between groups were performed using one-way analysis of variance (ANOVA). Results are expressed as mean ± standard error of the mean (SEM) and significance was set at *$p<0.05$, **$p<0.01$ and ***$p<0.001$.

CHAPTER 4 - RESULTS AND DISCUSSION

4.1. Design and cloning of sgRNAs targeting the human *ATXN2* gene

Despite the efforts from research groups in academia and other type of institutions to develop a permanent disease-modifying therapy, SCA2 treatment is limited to symptom alleviation. Arguably, the most promising therapy currently under development for SCA2 is the one that employs ASOs as a way to suppress ataxin-2 translation. This therapy consists in targeting the *ATXN2* mRNA with a complementary sequence of modified oligonucleotides, that will promote *ATXN2* mRNA degradation and therefore reduce the amount of mRNA available to the translated (Scoles et al 2017). Although the SCA2-127Q and BAC-72Q transgenic mice models treated with this ASO displayed significant improvement in motor function and a reduction of *ATXN2* expression, this therapy does not inhibit the transcription of potentially toxic CAG-expanded *ATXN2* RNA nor does it confer a permanent modification to protein translation (Martí 2016).

The CRISPR/Cas9 system offers the possibility of countering a genetic disease at the genomic level, the most upstream point of the pathological cascade, through gene-specific knockouts or replacement/repair of mutations. Gene editing can be regarded as an optimal therapeutical approach to treat genetic disorders such as SCA2. In fact, studies performed in Huntington's disease (Shin et al 2016) and SCA3 (Ouyang et al 2018), using CRISPR/Cas9-based therapies, reinforce the great potential this system has to treat polyQ disorders. Taking this into consideration, we sought to apply the CRISPR/Cas9 system to develop a putative therapeutic strategy for SCA2 that represented an alternative to mRNA-directed gene silencing (Matos et al 2018).

We decided to explore two different CRISPR/Cas9-based methods: one using a "traditional", catalytically active, Cas9 that would knock-out the *ATXN2* gene by introducing a mutation, and another employing a fusion protein composed of inactive dCas9 and KRAB, a transcription repressor, that would silence *ATXN2* at a pre-transcriptional level. To guide Cas9 to the intended target site, a sgRNA which we termed sg_ATXN2_Indel was designed to recognize an early site of *ATXN2* codifying region: a sequence in between the translation initiation codon and the beginning of the CAG repeats, in exon 1 (**Figure 7**). To guide the dCas9-KRAB fusion and maximize the efficiency of this strategy, according to Gilbert *et al.*, several sgRNAs (sg_ATXN2_Prom1 to 5) targeting different regions of the *ATXN2* promoter sequence were also designed (**Figure 7**) (Gilbert et al 2013).

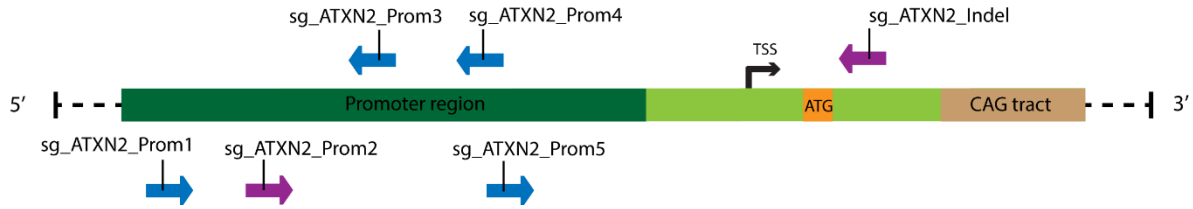

Figure 7- Location of the sgRNAs target sites in the *ATXN2* gene. The binding site for the sgRNA sg_ATXN2_Indel is located in the vicinity of the initiation codon and binds to the coding strand of *ATXN2*. sgRNAs targeting the promoter sg_ATXN2_Prom3 and sg_ATXN2_Prom4 also bind to the coding strand, while sgRNAs sg_ATXN2_Prom1, sg_ATXN2_Prom2 and sg_ATXN2_Prom5 are expected to bind to the template strand. Although 6 sgRNAs were designed, only sg_ATXN2_Prom2 and sg_ATXN2_Indel (purple arrows) were successfully cloned into the respective expression plasmid. The size of the annotated regions is not scaled.

The variable regions of the sgRNAs - the 20 nucleotide sequences that bind to their complementary DNA targets - were cloned in the form of double-stranded DNA oligonucleotides into expression plasmids that also comprised the respective Cas9 constructs. The quality and integrity of the resulting plasmids was assessed to ensure no significant structural changes occurred during the cloning process. To achieve this, the plasmids expressing Cas9 and sg_ATXN2_Indel (Cas9/Indel), and dCas9-KRAB and sg_ATXN2_Prom2 (dCas9-KRAB/Prom2), as well as their respective empty plasmids, LentiCRISPR v2 and dCas9-KRAB, were digested with NcoI restriction enzyme and the resulting fragments were separated and analyzed by agarose gel electrophoresis (**Figure 8**). The size of the DNA fragments resulting from the cloned plasmids digestion were compared to that of their backbones, so major changes to the plasmids structure during the cloning process would be identified. LentiCRISPR v2 was expected to generate four DNA fragments with, approximately, 6714, 4496, 3443 and 220 base pairs (bp), and dCas9-KRAB was expected to generate four fragments with sizes around 7588, 3648, 3443 and 220 bp.

Enzymatic restriction of LentiCRISPR v2 yielded four bands with sizes corresponding to what was expected, while Cas9/Indel displayed three bands with the same size as the control LentiCRISPR v2 plasmid, but did not present the fragment with approximately 4496 bp (**Figure 8**). Instead, Cas9/Indel yielded a fourth fragment with around 2600 bp. It was expected that an 1881 bp fragment of filler DNA would be removed from the region where the 20-nucleotide sgRNA sequence was introduced, and that region is contained in the original 4496 bp DNA fragment. The cloning process thus reduced the size of that fragment from 4496 bp to around 2636 bp, as expected.

dCas9-KRAB displayed three bands instead of the four bands that were expected, with corresponding sizes of around 7500, 3500 and 220 bp (**Figure 8**). However, the single band of around 3500 pb admittedly corresponds to two fragments, with 3648 and 3443 bp; their approximate size probably does not allow them to be detected as two distinct bands under this type of experimental setup. The fragments resulting from dCas9-KRAB/Prom2 enzymatic restriction correspond to the ones observed in its control, dCas9-KRAB (**Figure 8**). In this case, cloning was not expected to significantly alter the plasmid size, considering the small size of the filler region (16 bp).

Since the enzymatic restriction analysis revealed no major changes to the plasmids, such as large deletions or insertions, we considered that the cloned plasmids retained their integrity throughout the cloning process.

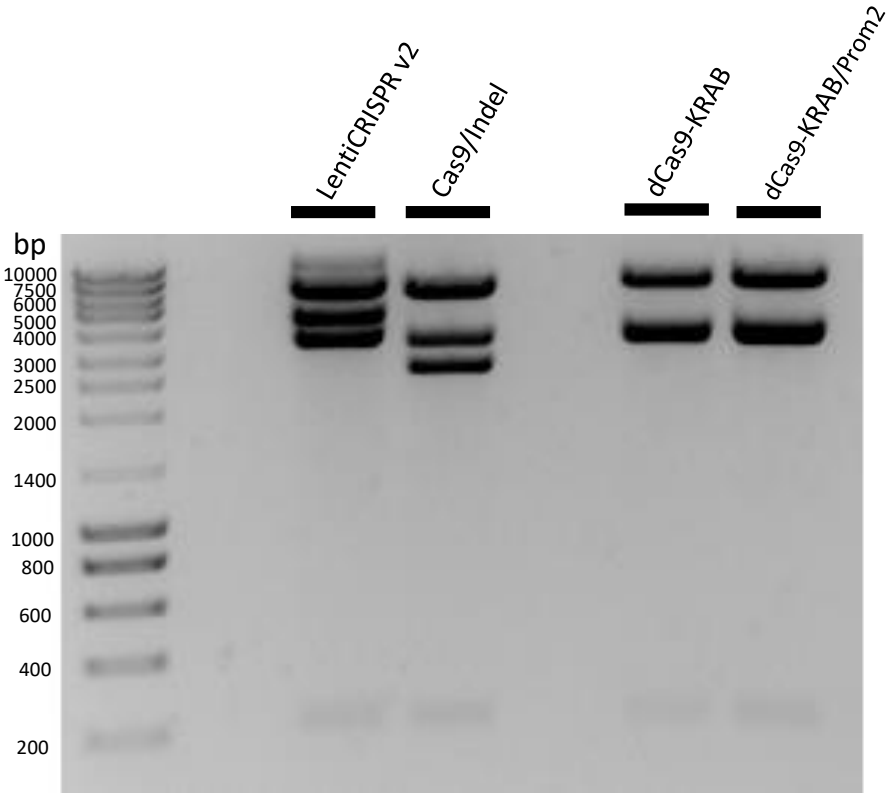

Figure 8- Restriction analysis of plasmids Cas9/Indel and dCas9-KRAB/Prom2. All plasmids were digested with NcoI and resulting fragments separated in a 1 % (w/v) agarose gel by electrophoresis. NZYDNA marker III standard was loaded alongside the plasmid samples. The integrity of the cloned plasmids was assessed by comparing the size of the DNA fragments resulting from enzymatic restriction analysis with the fragments of their respective expression plasmid backbones. LentiCRISPR v2 displayed four DNA fragments with approximately 67714, 4496, 3443 and 220 base pairs (bp), while Cas9/Indel displayed fragments with the same size apart from the 4496 bp fragment. Instead, Cas9/Indel displayed a DNA fragment with around 2600 bp, which was to be expected since a 1885 bp of filler DNA would be removed during the cloning process. Enzymatic digestion of dCas9-KRAB yielded three fragments instead of the expected four with 7588, 3648, 3443, and 220 bp. This can be explained with the approximate size of the fragments with 3648 and 3443 bp, that does not allow the two corresponding

bands to be distinguished. The fragments resulting from dCas9-KRAB/Prom2 correspond to the same observed in the corresponding expression plasmid backbone, KRAB_backbone. Together, this data indicated that the expression plasmids did not underwent noticeable alterations during cloning of the sgRNAs sequences.

To evaluate if the sgRNA sequences were correctly inserted into the sgRNA scaffold region of the plasmids, we performed Sanger sequencing using a primer targeting the U6 promotor, responsible for driving sgRNA expression in both plasmids.Sequencing analysis of the purified plasmidic DNAs showed that sg_ATXN2_Indel retained its sequence throughout the cloning process and was correctly introduced in the sgRNA scaffold from the plasmid expressing Cas9 (**Figure 9A**). As for the different sg_ATXN2_Prom sequences, Prom1, 3, 4 and 5 failed to be integrated into the dCas9-KRAB expression plasmid. Only sg_ATXN2_Prom2 was correctly introduced into the sgRNA scaffold region of the dCas9-KRAB expression plasmid and retained its full original sequence (**Figure 9B**). Together, the sequencing results obtained demonstrated that the insertion of designed sgRNAs sg_ATXN2_Indel and sg_ATXN2_Prom2 was successful, and that their sequences remained unaltered and complementary to the target site for which they were designed.

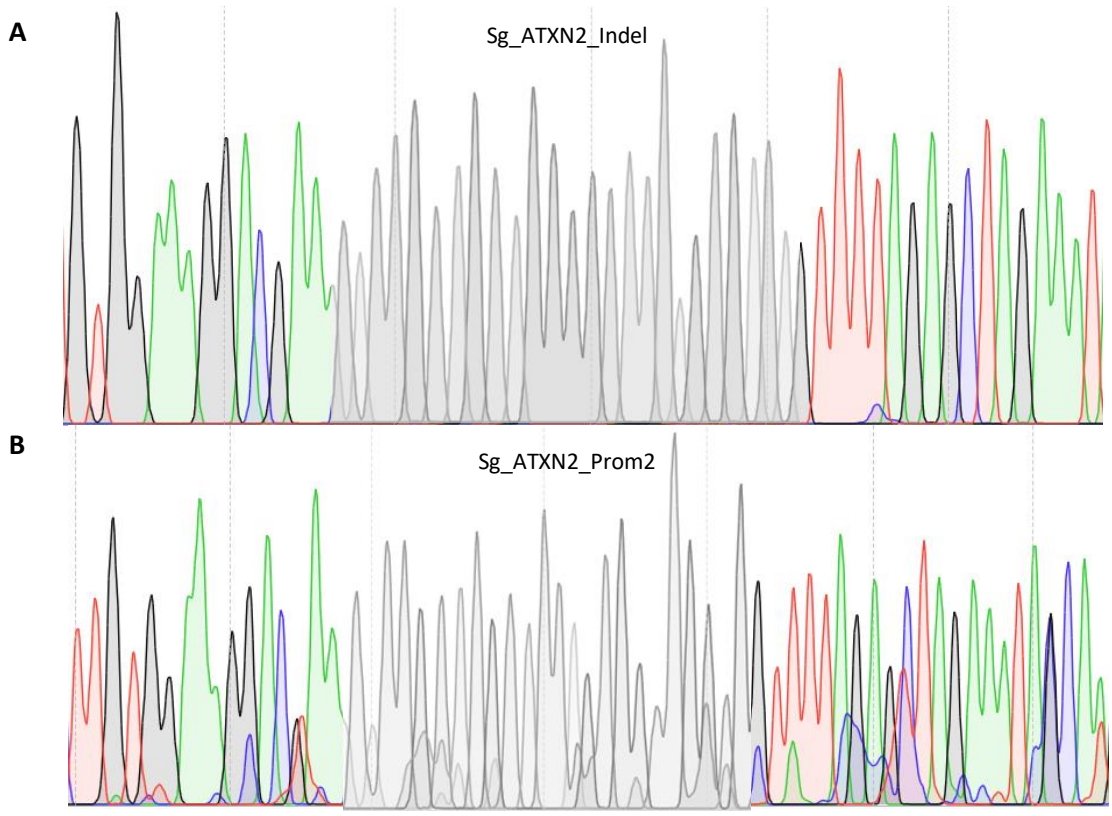

Figure 9- Sequencing chromatograms from the Cas9/Indel and dCas9-KRAB/Prom2 plasmids. Purified plasmid DNA was analyzed by Sanger sequencing, in order to confirm the introduction of the designed sgRNA sequences into the intended expression plasmids. A primer targeting the U6 promoter

present in both expression plasmids was used as the sequencing primer. Sequencing of the sgRNA insert region revealed that **(A)** the designed sg_ATXN2_Indel sgRNA was correctly inserted into the plasmid expressing Cas9, and that **(B)** the designed sg_ATXN2_Prom2 sgRNA was correctly inserted into the plasmid expressing dCas9-KRAB. The sequences of the sgRNAs is censured from the chromatograms, so as to remain undisclosed.

Together, the results from the enzymatic restriction analysis and Sanger sequencing indicate that the cloning of constructs Cas9/Indel and dCas9-KRAB/Prom2 was successful and did not compromise the overall integrity of the plasmids. Plasmid dCas9-KRAB/Prom2 sequencing did not yield a completely clear chromatogram, but this was attributed to defects in the sequencing process, which may have unspecifically amplified sequences other than the one intended. The two engineered constructs were thus deemed viable to be used as the expression systems of the two CRISPR-based silencing strategies that the current project aimed to develop. The ability of these constructs to elicit repression of human *ATXN2* expression was tested in transfected human cell line cultures.

4.2. Silencing *ATXN2* through the introduction of a premature STOP codon

Cas9 is an RNA-programmable endonuclease, capable of introducing targeted DNA DSBs after being guided by a sgRNA molecule that is complementary to the target site. The repair of DSBs in mammalian DNA, in the absence of a repair template, is performed by NHEJ, a pathway prone to errors that often result in the generation of indel mutations at the repair site (Cong et al 2013, Guo et al 2018, Maeder & Gersbach 2016). Exploiting this mechanisms, targeted DSBs mediated by Cas9 can be used to stimulate NHEJ at a region of interest. The indels that are formed disrupt gene expression when the corresponding insertions or deletions lead to the introduction of premature STOP codons due to DNA reading frame shifts (Cong et al 2013).

We sought to develop a putative therapeutic strategy for SCA2 based on impairing formation of the ataxin-2 protein and thus preventing its toxicity. One of the approaches that were envisioned relied on the generation of a very early premature STOP codon in *ATXN2*, possibly upstream of the CAG repeats. This mutation would allow long-term blocking of the gene, at the DNA level, ensuring that only a very limited number of *ATXN2* codons would be translated and preventing the polyQ tract from being assembled. To accomplish this, we employed Cas9 guided by the sg_ATXN2_Indel sgRNA to introduce a DSB in between the

translation initiation codon and the beginning of the CAG repeats. The introduction of this DSB would stimulate NHEJ to introduce an indel at the repaired site and lead to the formation of a premature STOP codon downstream of that region (**Figure 10**).

This Cas9-based strategy was tested in HEK 293T cells, a human cell line that multiplies rapidly, and that is easy to maintain and transfect. Cultures were transfected with the Cas9/Indel plasmid, codifying both Cas9 and the sg_ATXN2_Indel sgRNA, as well as a puromycin resistance gene. 48 h after transfection, cultured cells were treated with 10 µg/mL puromycin, so that transfected cells could be positively selected. As a cut-inefficient control for Cas9/Indel we used the Cas9/LacZ plasmid, codifying both Cas9 and a sgRNA sequence targeting *Escherichia coli* LacZ gene, that has been validated as having no targets in human cells (Platt et al 2014).

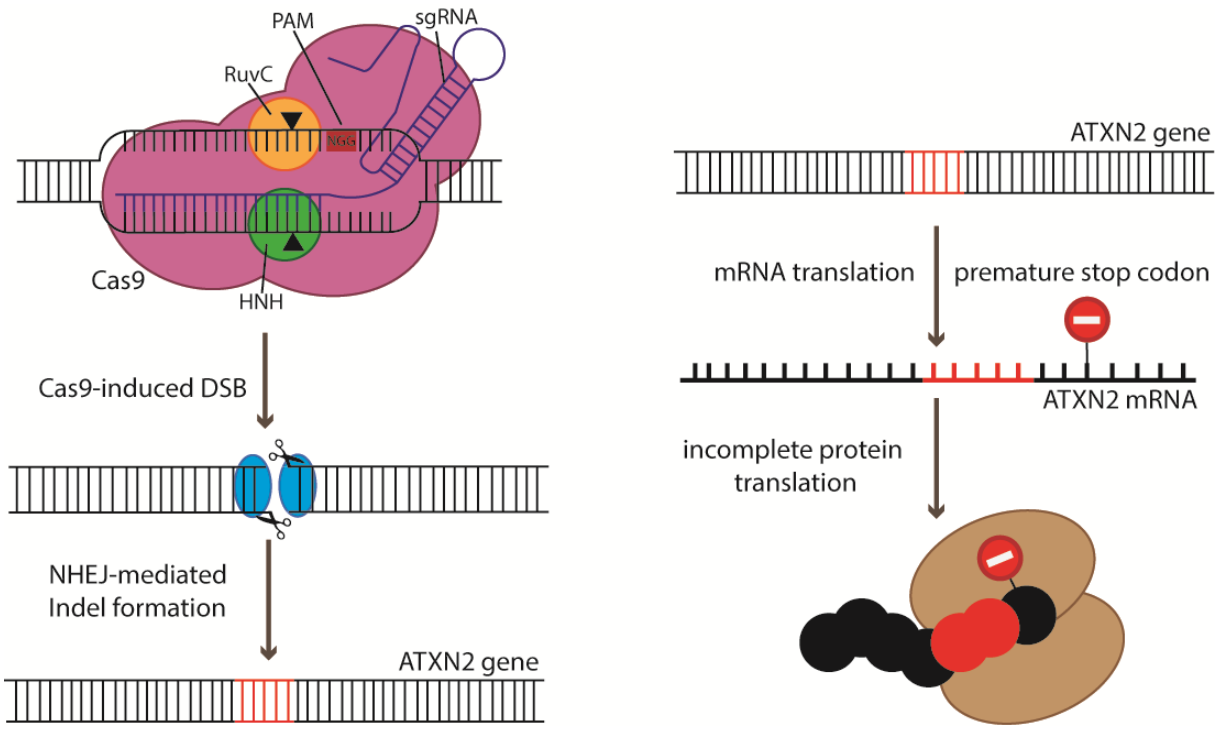

Figure 10- Formation of a premature start codon in *ATXN2* via Cas9-induced DSB. The Cas9 ribonucleoprotein complex is constituted by the Cas9 protein (purple) and the sgRNA molecule (blue). Cas9 has two catalytic domains, RuvC (orange) and HNH (green), which induce DNA breaks (▼) 3 base pairs upstream of the PAM (5'-NGG-3') (red box). Each DNA strand will be nicked by one of Cas9 catalytic domains, giving rise to a DSB. NHEJ-mediated repair of the Cas9-induced DSB can culminate in the formation of an indel mutation, that in turn may lead to the generation a premature stop codon (no entry sign) in the *ATXN2* gene. Translation of the mRNA molecule containing the introduced premature stop codon will result in disturbed ataxin-2 translation, leading to the production of a drastically-truncated protein with a very limited number of amino acids and possibly lacking the polyQ region tract.

Since our Cas9/Indel strategy aimed at editing the *ATXN2* gene sequence permanently and impair ataxin-2 translation, we evaluated whether expression of Cas9/Indel in HEK 293T decreased expression of the endogenous *ATXN2* gene. For that, we assessed ATXN2 mRNA and protein levels. Additionally, to determine if indel mutations had in fact been inserted at the target region, a fragment of the *ATXN2* gene containing that region was sequenced.

To assess whether Cas9/Indel expression affected *ATXN2* mRNA levels, total RNA extracts from transfected HEK 293T were subjected to a qPCR analysis. As a control for *ATXN2* mRNA expression we used the Glyceraldehyde 3-phosphate dehydrogenase *(GAPDH*) gene (the same gene used as control in Western blot experiments). Considering that the indel-based knock down strategy is not expected to interfere with ATXN2 transcription directly, a major alteration to the mRNA levels was not anticipated upon HEK 293T cells transfection with Cas9/Indel, in comparison to the controls. Accordingly, qPCR results indicated that expression levels of *ATXN2* mRNA in HEK 293T transfected with Cas9/Indel did not display a statistically significant difference from those of cell transfected with Cas9/LacZ or the NT condition (**Figure 11**). Importantly, however, results obtained in the different experimental replicates of the transfected conditions were highly variable (as represented by the large SEM bars).

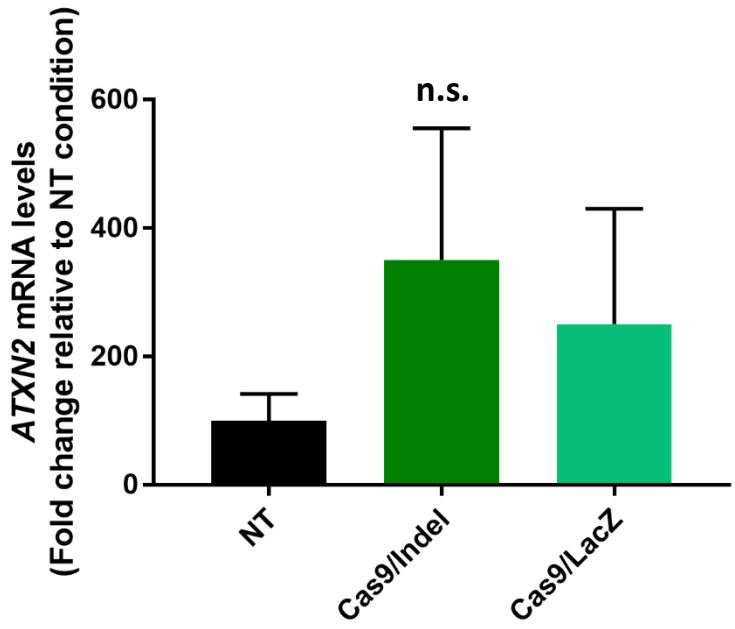

Figure 11- *ATXN2* mRNA levels in HEK 293T cells expressing Cas9/Indel and Cas9/LacZ. qPCR was performed with cDNA converted from total RNA extracts of HEK 293T cells transfected with plasmids coding for Cas9 and a sgRNA targeting the region upstream of the CAG tract of *ATXN2* or, as a control, with plasmids coding for Cas9 and a sgRNA targeting LacZ, a gene absent from eukaryotic cells. *GAPDH* was used as a control gene for normalization. qPCR analysis revealed no difference in *ATXN2* mRNA levels between cells transfected with Cas9/Indel and the respective controls (Non-transfected, n = 3; Cas9/Indel, n =3; Cas9/LacZ, n =3). Data is presented as mean ± SEM. Non-significant – n.s.: $p > 0,05$ unpaired Student´s t-test.

In order to investigate if there was a difference in ataxin-2 protein levels between cells expressing Cas9/Indel or the control Cas9/LacZ, lysates from the transfected HEK 293T cells were subjected to a Western blot analysis. Membranes were probed with an anti-ataxin-2 antibody, and as a normalization control probed with an anti-GAPDH antibody (**Figure 12A**).

Since our strategy aimed at disrupting ataxin-2 protein translation, we expected a reduction of ataxin-2 levels in cells transfected with Cas9/Indel, when compared to control conditions. Protein level assessments revealed that cells transfected with Cas9/Indel displayed a tendency for a decrease in ataxin-2 protein levels. However, the difference was not statistically significant, when comparing Cas9/Indel to Cas9/LacZ, or to the untransfected condition (**Figure 12B**). This may be attributed to the high values of standard error, resulting from the variability observed between replicates.

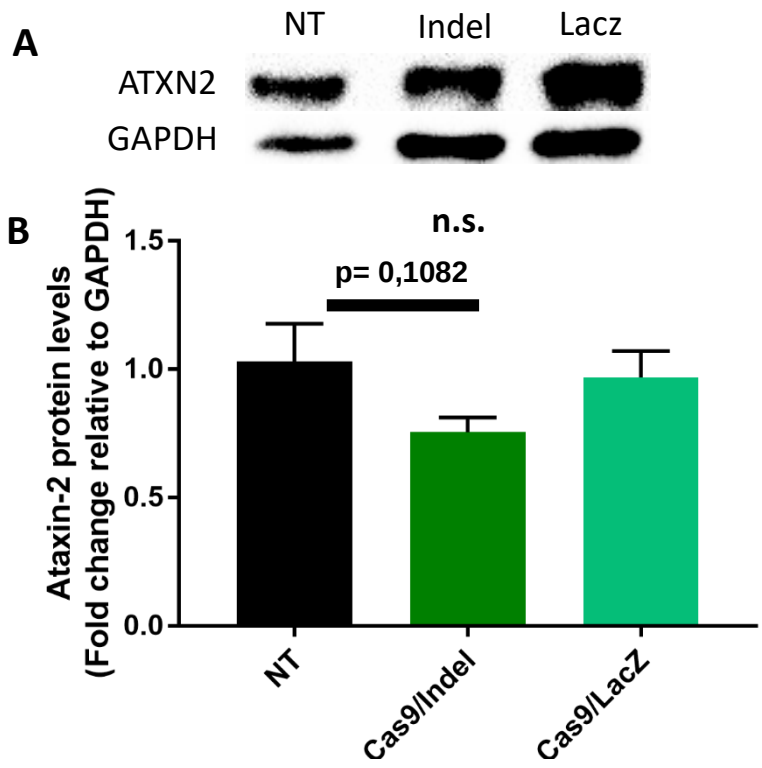

Figure 12 – Ataxin-2 protein levels in HEK 293T cells expressing Cas9/Indel or Cas9/LacZ. (A) Anti-ataxin-2 and anti-GAPDH antibody Western blot staining were performed on total protein lysates from HEK 293T cells transfected with plasmids coding for Cas9 and the sgRNA targeting the region upstream the CAG tract of *ATXN2* or, as a control, with plasmids coding for Cas9 and a sgRNA targeting LacZ, a structural gene absent from eukaryotic cells. **(B)** Comparison between cells transfected with Cas/Indel and the respective controls revealed no significative differences regarding protein levels. All the values are normalized to the respective GAPDH levels (Non-transfected (NT), n = 3; Cas9/Indel, n =4; Cas9/LacZ, n =4). Data is presented as mean ± SEM. Non-significant – n.s.: $p > 0,05$ unpaired Student´s t-test.

To determine if the DSB generated by our gene editing tool led to the introduction of an indel mutation and the consequent appearance of a premature STOP codon in the downstream vicinity, a fragment of the *ATXN2* gene from HEK 293T cells transfected with Cas9/Indel or Cas9/LacZ was sequenced. Genomic DNA was extracted and an *ATXN2* gene region comprising the site targeted by Cas9/Indel was amplified by PCR, using a pair of custom primers: the forward primer targeted a site upstream of the *ATXN2* start codon, and the reverse primer targeted a region downstream of the site where the DSB would be inserted. PCR products were then separated by agarose gel electrophoresis and the amplicon of interest (950 bp) was isolated from the gel and submitted to Sanger sequencing (**Figure 13**). Importantly, agarose gel separation revealed that several unspecific products were also generated, which may indicate that the primers used were not sufficiently specific, even though the amplicon of interest corresponded to the most intense band.

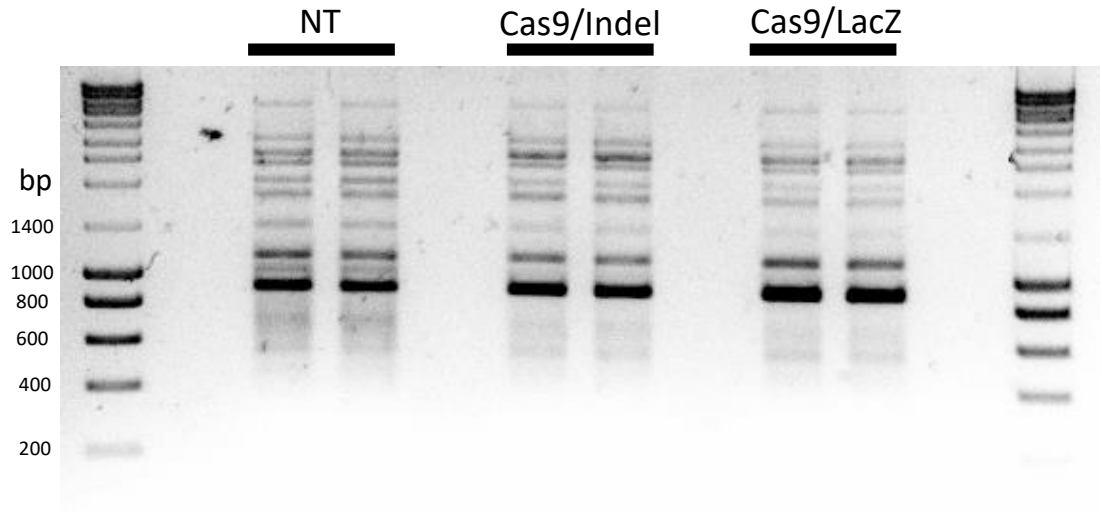

Figure 13- PCR products of genomic DNA samples from HEK 293T cells transfected with Cas9/Indel. Genomic DNA from HEK 293T cells transfected with Cas9/Indel was purified and amplified by PCR using a pair of custom primers, targeting regions upstream and downstream of the expected DSB site of Cas9. Each PCR product was loaded into 2 lanes of a 1,5 % (w/v) agarose gel and separated by electrophoresis. The PCR product of interest was calculated with SnapGene to be around 950 bp. Agarose gel electrophoresis analysis of the PCR products revealed an amplicon with increased staining of approximately 950 bp, which corresponded to the amplicon of interest. Many unspecific products were also detected.

Analysis of the sequencing chromatograms demonstrated that the amplicon sequence from NT (**Figure 14 A**), Cas9/Indel (**Figure 14 B**) and Cas9/LacZ (**Figure 14 C**) conditions displayed a significant amount of background "noise", that is, several nucleotide spikes overlap in the same position, making for an unreliable sequence output. This way, we were unable to

analyze the sequences for the presence of indels. The excessive noise suggests that more than one amplicon may have been isolated form the agarose gel and then sequenced.

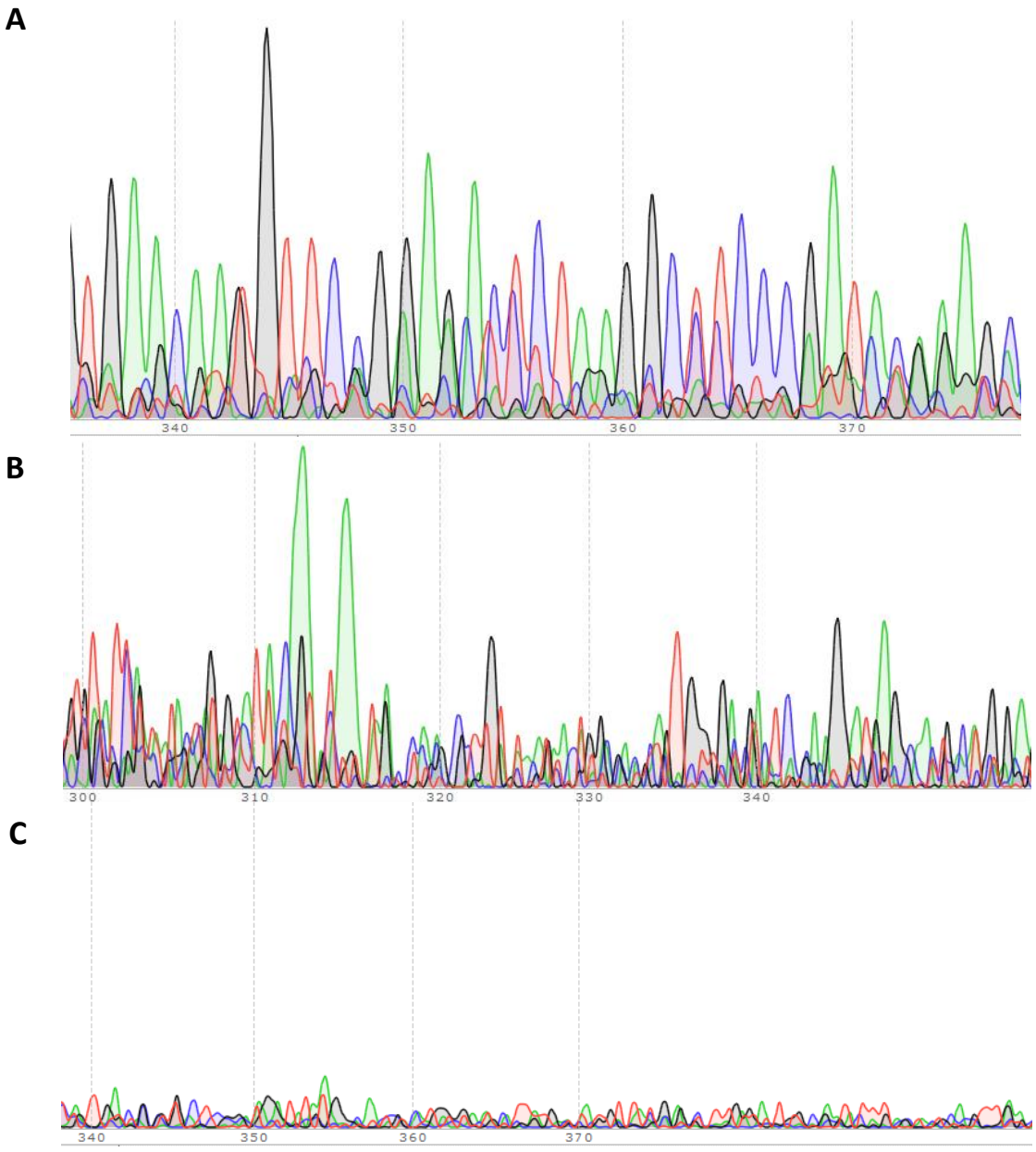

Figure 14- Sequencing chromatograms of *ATXN2* PCR amplicons containing the Cas9/Indel target site. The 950 bp amplicon of interest was isolated in each condition from the agarose gel represented in Figure 13 and submitted to Sanger sequencing. Chromatograms generated by conditions NT**(A)**, Cas9/Indel **(B)** and Cas9/LacZ **(C)** revealed several nucleotide spikes overlapping in the same positions, all throughout the sequences. This suggests that unspecific PCR products may have been simultaneously sequenced during the process.

Considering that this indel-based strategy would theoretically impair the translation of a full-length ataxin-2 protein, a more prominent decrease of an ataxin-2 protein product with 144 kDa was expected in cells transfected with Cas9/Indel. The modest results obtained may be explained with a dilution effect on the premature STOP codons that may be formed by indel introduction, at the intended position. The indels that are generated by NHEJ are variable from cell to cell; accordingly, while in a proportion of the transfected cell population a premature STOP codon may be generated, in another fraction of the cells nucleotide additions or deletions will not alter the reading frame (this happens if that number of nucleotides is a multiple of 3). Additionally, positive selection of transfected cells using puromycin does not depend on indel generation at the *loci* of interest, and thus unaltered *ATXN2* alleles may remain in the transfected cell population. Considering that analysis of protein levels concerns the whole transfected cell population, the effect of our strategy would be diluted as a result of the above circumstances, and that is possibly the reason why the differences in ataxin-2 protein levels did not reach statistical significance.

It is possible to deconvolute and dissect the chromatograms resulting from an indel-bearing sequence through a process termed Tracking of Indels by Decomposition (TIDE) analysis (Brinkman et al 2014). Sanger sequencing results from the amplicons corresponding to the *ATXN2* region targeted by sg_ATXN2_Indel were intended to be subjected to this TIDE analysis, which would allow the detection of the presence of the indels and analysis of their constitution. However, the chromatograms obtained failed to enlighten whether an indels was efficiently generated at the DSB site. The continuous overlap of nucleotide spikes, which impairs the intended analysis of the chromatograms, suggested the presence of one or more amplicons in the agarose gel band from which DNA was isolated for sequencing. The primers used in the PCR procedure may not have been sufficiently specific, amplifying regions other than the one intended.

4.3. Silencing ATXN2 through dCas9-KRAB-mediated transcriptional repression

CRISPR/Cas9 is an inexpensive, relatively easy to use, and versatile gene editing system that can theoretically be targeted to any genetic *locus*. This facts made this system into an unprecedented tool to drive gene editing in the human genome (Aryal et al 2018). Nevertheless, CRISPR/Cas9 is still a fallible, prone to error, system, which may lead to off-target insertion of DSBs. Also, gene editing with conventional CRISPR/Cas9 relies on endogenous DNA repair mechanisms to insert modifications into the target site, which are independent from the tool itself and can have variable outcomes depending on how DNA is repaired (Zhang et al 2015).

CRISPR interference is proposed as a potentially safer gene silencing tool comparing to nuclease activity-based gene editing platforms, since CRISPRi does not rely on DSBs or endogenous DNA repair mechanisms. This alternative CRISPR/Cas-based system utilizes a catalytically dead Cas9 (dCas9) protein, fused to a transcription repressor such as the Krüppel-associated box (KRAB) domain (Gilbert et al 2013). dCas9-KRAB may enable repressing *ATXN2* expression at the pre-transcriptional level and therefore prevent the production not only of the pathogenic ataxin-2 protein, but also of the toxic CAG-expanded *ATXN2* mRNA, that may contribute to SCA2 pathogenesis (Martí 2016).

We sought to develop a CRISPRi-based therapeutical approach to SCA2, that would silence *ATXN2* by hindering the binding of the transcriptional machinery to the promoter. To achieve this, we employed a dCas9-KRAB fusion protein guided by the sgRNA termed sg_ATXN2_Prom2, that would target the fusion to the promoter of the *ATXN2* gene. Once the CRISPRi complex is bound to the promoter, KRAB would be expected to recruit chromatin-modifying machinery that would promote the condensation of the chromatin region where *ATXN2* is located (Gilbert et al 2013). Due to the modification in the chromatin tridimensional structure, transcription factors would not be able to bind to the promoter region of *ATXN2* (**Figure 15**). This would ultimately repress gene expression, resulting in a decrease of *ATXN2* mRNA and protein levels.

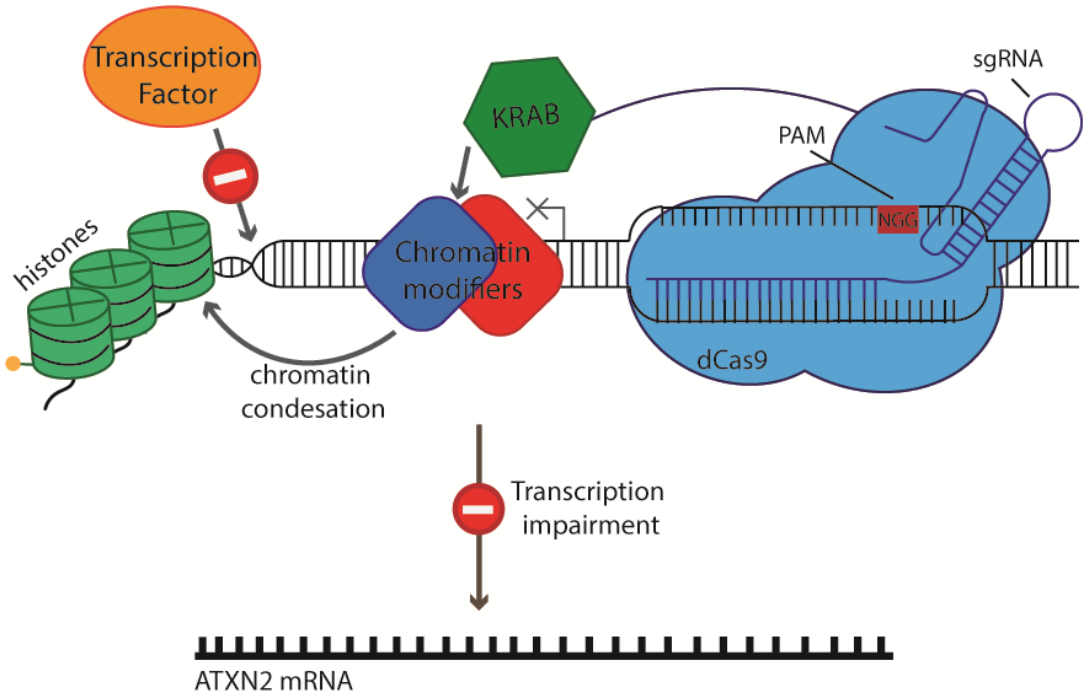

Figure 15 - Inhibiting *ATXN2* transcription via CRISPRi. dCas9 (light blue) has both catalytic domains mutated, impairing the ability to induce DSBs. When fused to a transcription repressor (KRAB), which promotes the recruitment of chromatin-modifying machinery, the dCas9-KRAB fusion is capable of silencing gene expression. Chromatin modifiers will alter the tridimensional structure of the chromatin, so that transcription factors will not be able to bind to the promoter. This will therefore repress gene transcription. If directed to the *ATXN2* gene promoter, this strategy is expected to result in the impairment of the production of the potentially toxic *ATXN2* mRNA and of the pathogenic ataxin-2 protein.

In order to evaluate the efficacy of our approach, and identically to what was done for the Cas9-based strategy (section 3.2), HEK 293T cells were transfected with the dCas9-KRAB/Prom2 plasmid, expressing both dCas9-KRAB and the sg_ATXN2_Prom2, as well as a puromycin resistance cassette. 48 h after transfection, cultured cells were treated with 10 µg/mL puromycin, so that the transfected cells could be selected. Since this dCas9-KRAB-based strategy is expected to repress *ATXN2* expression at a pre-transcriptional level, we assessed whether ATXN2 mRNA and protein levels were downregulated after 72 h after transfection. As a control for dCas9-KRAB/Prom2, we used a dCas9-KRAB construct with an incomplete sgRNA inserted into the sgRNA scaffold (the empty dCas9-KRAB vector; dCas9-KRAB/Empty).

With the constraint in *ATXN2* mRNA transcription, we expected a reduction in the amount of *ATXN2* mRNA molecules available to be translated into protein, thus decreasing ataxin-2 levels. In order to investigate the effects of dCas9-KRAB/Prom2 on *ATXN2* expression in HEK 293T cells, *ATXN2* mRNA levels were assessed by qPCR. Contrary to what we had

anticipated, *ATXN2* mRNA levels in the different experimental conditions did not show any statistically significant differences (**Figure 16**).

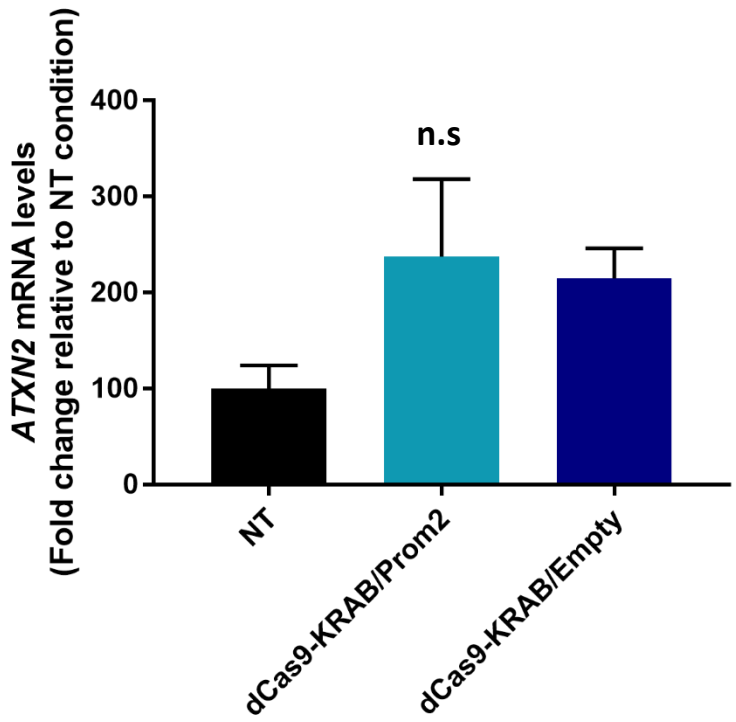

Figure 16- *ATXN2* mRNA levels in HEK 293T cells expressing dCas9-KRAB/Prom2 or dCas9-KRAB/Empty. qPCR was performed with cDNA converted from total RNA extracts of HEK 293T cells transfected with plasmids coding for dCas9-KRAB and the sgRNA targeting *ATXN2* promoter region (dCas9-KRAB/Prom2) or, as control, with plasmids coding for dCas9-KRAB with an incomplete sgRNA inserted into the plasmid (dCas9-KRAB/Empty). *GAPDH* was used as a control gene for normalization. qPCR analysis revealed no statistically significant differences in *ATXN2* mRNA levels between cells transfected with dCas9-KRAB/Prom2 and the respective controls. All the values are normalized to the respective *GAPDH* levels (Non-transfected, n = 3; dCas9-KRAB/Prom2, n =3; dCas9-KRAB/Empty, n =3). Data is presented as mean ± SEM. n.s.: p>0,05 unpaired Student´s t test.

To investigate whether ataxin-2 protein levels were reduced, we performed a Western blot analysis of cell lysates from HEK 293T cell cultures transfected with dCas9-KRAB/Prom2. The Western blot membrane was probed with an anti-ataxin-2 antibody, and as a normalization control probed with an anti-GAPDH antibody (**Figure 17A**). Results revealed that HEK 293T cells transfected with the dCas9-KRAB/Prom2 silencing tool displayed a tendency for a decrease of ataxin-2 levels, when comparing to non-transfected cells (p value = 0,0534). No differences were observed when comparing to the control dCas9-KRAB/Empty condition, admittedly due to the large SEM value (**Figure 17B**).

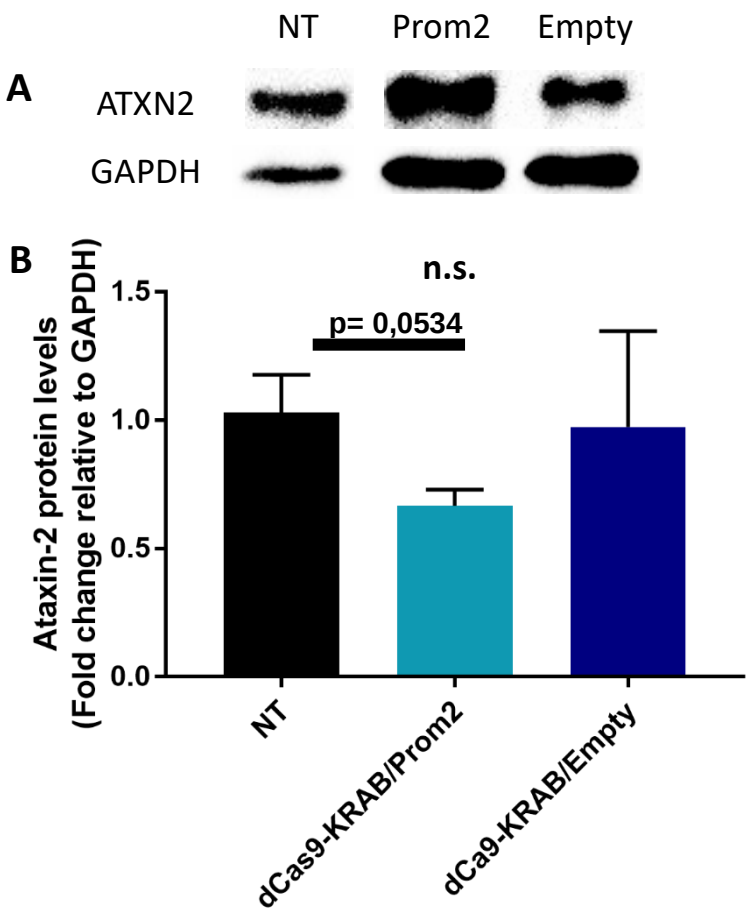

Figure 17 - Ataxin-2 protein levels in HEK 293T cells expressing dCas9-KRAB/Prom2 or dCas9-KRAB/Empty. (A) Anti-ataxin-2 and Anti-GAPDH antibody Western blot stainings were performed on total protein extracts from HEK 293T cells transfected with plasmids coding for dCas9-KRAB and the sgRNA targeting *ATXN2* promoter region (dCas9-KRAB/Prom2) or, as control, with plasmids coding for dCas9-KRAB with an sgRNA inserted into the plasmid (dCas9-KRAB/Empty). **(B)** Comparison between cells transfected with dCas9-KRAB/Prom2 and the respective controls revealed no significative differences regarding protein levels. However, a tendency for a decrease of ataxin-2 levels was observed in cells treated with dCas9-KRAB/Prom2, when comparing to non-transfected cells (p value = 0,0534; NT, n = 3; dCas9-KRAB/Prom2, n =4; dCas9-KRAB/Empty, n =4). Data is presented as mean ± SEM. n.s.: p > 0,05 unpaired Student´s t-test.

The considerable variation of the qPCR results that existed between different experimental replicates, and which resulted in high SEM values, suggests that there were problems during the qPCR procedure. The problems encountered were possibly due to the *ATXN2* primers used, considering that negative controls conditions containing the *ATXN2* qPCR primers generated significant qPCR signals even in the absence of template cDNA. Interestingly, after testing all qPCR reagents for DNA or reverse transcriptase contamination, no reagent proved to be contaminated with either substance. The *ATXN2* primers used also displayed amplification efficiency values above the maximum recommended, which influenced

the *ATXN2* expression standards used for the calibration curve. Together these setbacks may have ultimately affected the expression values that were determined. In order to eliminate the problems encountered and appropriately assess the efficiency of the developed gene silencing tool, different sets of *ATXN2* primers should be tested and optimized.

The reduction observed in ataxin-2 protein levels was admittedly modest and did not reach statistical significance. The 72 hours incubation period during which cells were maintained after transfection may not have been enough for the effect of our tool to be sufficiently noticeable. It is also possible that the region of the *ATXN2* promoter targeted by sg_ATXN2_Prom2 may not represent an optimal target for dCas9-KRAB. That way, dCas9-KRAB/Prom2 may not completely block transcription factor binding to the promoter and therefore transcription is not adequately inhibited. Taken that the exact region of the *ATXN2* promotor where the transcription factors and transcription modulators bind has not been described, it may be worthy to test the effect of other sgRNAs targeting the promoter. These other sgRNAs may direct dCas9-KRAB to positions that are more crucial to ATXN2 transcription initiation. In fact, as mentioned before, four other sgRNA sequences targeting the ATXN2 promoter were designed as part of the current project, but were not successfully cloned into the plasmid expressing dCas9-KRAB.

CHAPTER 5 - CONCLUSION AND FUTURE PERSPECTIVES

Currently, there is no effective curative treatment for SCA2, so patients have to rely solely on symptomatic control. Moreover, the cellular and molecular mechanisms underlying this polyQ disorder are still poorly understood, hindering advances in the development of a therapy. Future research will help draw the full picture of the disease's pathogenesis. Although there is still much to discover about SCA2, it has been described that an abnormal expansion beyond 30 CAG repeats in the exon 1 of *ATXN2* causes the disease, and culminates in the production of an aberrant protein with a polyQ-expanded tract, admitted to play a central role in SCA2 pathogenesis.

The work hereby presented had the objective of exploring the therapeutic potential of gene editing/gene silencing strategies demonstrated in other polyQ disorders, to develop a therapeutic approach to SCA2. Two routes to *ATXN2* gene silencing based on the CRISPR/Cas9 gene editing system were investigated: one based on indel-derived gene blocking, and another based on CRISPRi pre-transcriptional silencing. Although results did not support the efficacy of the indel-based strategy, cells submitted to the CRISPR interference approach demonstrated a tendency for ataxin-2 protein level reduction. Further improvements to our approaches will be necessary to clearly attest their potential. However, the results obtained during this yearlong work suggest that CRISPR/Cas9-based tools may represent a pathway for silencing the pathogenic *ATXN2* and treat SCA2 in the future.

In the section of this work regarding the indel-based strategy, we intended to develop a molecular tools capable of generating a premature STOP codon in the exon 1 of the human *ATXN2* gene. This gene edition would result in the incomplete translation of the ataxin-2 protein and ultimately repress ataxin-2 expression, possibly halting SCA2 pathogenesis in a disease context. With the aim of impairing ataxin-2 translation in HEK 293T human cell line, we used Cas9 guided by a sgRNA that was complementary to a region in the vicinity of the *ATXN2* START codon, prior to the beginning of the CAG repeats. Directing Cas9 to that region of *ATXN2* would result in the introduction of a DSB, that would be re-ligated through NHEJ. This DSB repair mechanism NHEJ often produces indels at the repair site, which may cause a premature stop codon to be generated. While this therapy was intended to impair ataxin-2 translation, Western blot results revealed that the levels of ataxin-2 were not significantly altered in cells transfected with the plasmid encoding the designed tools. Additionally, despite the efforts, we were not able to confirm that indel mutations or premature STOP codons were generated in the targeted region of *ATXN2*.

Together, the results obtained with the first approach to *ATXN2* silencing are insufficient to draw any conclusion regarding the efficiency of the gene editing tools designed. There were several practical complications associated with both the qPCR and the sequencing procedures. Considering the tendency for ataxin-2 protein levels reduction observed in the Western blot analysis, it is possible that a refinement of our experimental approaches will nonetheless prove the efficacy of this strategy. Increasing the number of experimental replicates may help counterbalance the high variability of the results obtained in different experiments. Also, sequence analysis for the detection of indels may be more successful if the PCR procedure ensured only the amplification of the intended ATXN2 gene region. This way, no contaminating amplicons would interfere with the sequencing process. Several pairs of PCR primers could be tested in their ability to amplify the edited segment of genomic DNA, and then the whole procedure should be optimized in order to produce the intended amplicon with the highest degree of specificity possible.

Another way of taking advantage of the CRISPR/Cas9 system ability to induce DNA DSB and thereby silence ataxin-2 would be to use the system to remove the START codon from the *ATXN2* altogether. To achieve this, contrary to what was performed in the current work, a pair of sgRNAs would be used, one sgRNA targeting a region upstream of the *ATXN2* START codon and another targeting a region downstream. The introduction of two Cas9-induced DSBs flaking both ends of the START codon would result in its removal, and posterior re-ligation of the DSB via NHEJ Without a START codon, ataxin-2 translation would be completely repressed. This approach would not rely on the stochasticity of the NHEJ pathway to produce a premature STOP codon, and this would possibly contribute to the ability to knock-out *ATXN2* in a larger population of cells (Giri et al 2019).

In the second part of this project, which regarded the CRISPRi-based strategy, we aimed at silencing the human *ATXN2* gene expression at a pre-transcriptional level, also in HEK 293T cells. With this strategy we sought to act at the most upstream point of gene expression by hindering the binding of transcription factors to the *ATXN2* promoter. In order to archive this, we used a catalytically dead Cas9, fused to a KRAB repressor domain (dCas9-KRAB), guided by a sgRNA targeting the promoter of the *ATXN2* gene. As a result, *ATXN2* mRNA would not be transcribed and *ATXN2* expression would be repressed. HEK 293T cells transfected with the construct codifying our silencing tool revealed a tendency for reduced ataxin-2 protein levels, as analyzed by Western blot. Contrastingly, qPCR results showed that *ATXN2* mRNA levels were not significantly altered by our silencing approach. We consider that these different

outcomes were due to problems encountered with the qPCR *ATXN2* primers, which most likely resulted in qPCR not relaying results that represent the true effect of our therapy in HEK 293T cells. The same may have been true for the qPCR analysis used in the first strategy tested.

Although the results obtained in the Western blot experiments seem to suggest that our silencing tool is able to repress *ATXN2* expression at some level, we cannot be certain until the experiments are further replicated and corroborated by viable qPCR experiments that demonstrate a reduction in *ATXN2* mRNA levels. Taken that our gene silencing tool does not produce permanent therapeutic effects, the determination of the ideal dose of DNA to be transfected, for maximum effect and duration, would also help produce a more pronounced and durable reduction.

The four sgRNA sequences (sg_ATXN2_Prom1, 3, 4 and 5) targeting the *ATXN2* promoter that were designed but were not successfully cloned into the plasmid expressing dCas9-KRAB, have been efficaciously inserted after the completion of the work hereby described. In the near future, we plan to determine which of the designed sgRNAs targeting different regions of the *ATXN2* promoter elicits a more pronounced *ATXN2* repression.

Considering that HEK 293T cells multiply rapidly and divide continuously, plasmids transfected into the cells would eventually be diluted in the daughter cells population, leading to a ever-decreasing expression of Cas9/Indel and dCas9-KRAB/Prom2 in the transfected cells. This is a possible additional explanation for the modest effects of the developed tools. As an alternative to counteract this expression problem in the future, we plan to employ lentiviral vectors carrying Cas9/Indel or dCas9-KRAB/Prom2 to transduce cells *in vitro*, and thus confer a stable and durable expression of the constructs in the cultured cells. This long-term strategy may potentiate the putative silencing effects of our tools and allow for a clear detection of those effects.

In case our approaches are efficiently demonstrated to significantly repress *ATXN2* expression in the human cell line model used in the current project, they will be tested in primary cell cultures retrieved from SCA2 patients, and then in patient iPSC-derived neurons. This will allow us to evaluate the effects of Cas9/Indel and dCas9-KRAB/Prom2 derived *ATXN2* silencing in neurons with the *ATXN2* pathological mutation. Posteriorly, the same tools will be tested in animal models, namely a mouse model with a SCA2 phenotype currently under development at our lab. This will enable us to assess the putative beneficial effect of our therapies on SCA2-related phenotypic traits and neuroanatomical features *in vivo.*

The strategies that were designed are admittedly unable to distinguish between non-expanded and expanded ATXN2 alleles. Although it has been shown that allele-unspecific silencing of CAG-bearing genes may have beneficial therapeutic effects in the context of polyQ diseases (Alves et al 2010), it would be of great value to evaluate the effects of CRISPR/Cas9-based silencing strategies specifically targeting mutant *ATXN2*, containing CAG repeats beyond the pathological range. Designing sgRNA sequences that specifically target pathogenic alleles is not always possible, but future efforts will also be directed to this objective.

Further studies and repetition of some of the experiments performed in this work will be necessary to establish and validate the potential of both the indel and the CRISPRi-based strategies as putative approaches to *ATXN2* silencing and SCA2 therapy. The current work demonstrated that the versatility of the CRISPR/Cas9 gene editing system allows the design of distinct gene silencing strategies. The fact that gene silencing may be enacted through a diverse set of mechanisms encourages exploration of any number of possible routes, and their respective advantages and disadvantages. With tools such as the ones we aim to develop in their arsenal, the scientific community will hopefully be closer than ever to engineer a curative treatment for SCA2, as well as other disorders with a genetic component to their pathophysiology.

References

Adli M. 2018. The CRISPR tool kit for genome editing and beyond. *Nat Commun* 9: 1911

Albrecht M, Golatta M, Wullner U, Lengauer T. 2004. Structural and functional analysis of ataxin-2 and ataxin-3. *European journal of biochemistry* 271: 3155-70

Alves S, Nascimento-Ferreira I, Dufour N, Hassig R, Auregan G, et al. 2010. Silencing ataxin-3 mitigates degeneration in a rat model of Machado–Joseph disease: no role for wild-type ataxin-3? *Human Molecular Genetics* 19: 2380-94

Anzalone AV, Randolph PB, Davis JR, Sousa AA, Koblan LW, et al. 2019. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 576: 149-57

Arnould S, Perez C, Cabaniols JP, Smith J, Gouble A, et al. 2007. Engineered I-CreI derivatives cleaving sequences from the human XPC gene can induce highly efficient gene correction in mammalian cells. *J Mol Biol* 371: 49-65

Aronin N, Kim M, Laforet G, DiFiglia M. 1999. Are there multiple pathways in the pathogenesis of Huntington's disease? *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 354: 995-1003

Aryal NK, Wasylishen AR, Lozano G. 2018. CRISPR/Cas9 can mediate high-efficiency off-target mutations in mice in vivo. *Cell death & disease* 9: 1099

Asada A, Yamazaki R, Kino Y, Saito T, Kimura T, et al. 2014. Cyclin-dependent kinase 5 phosphorylates and induces the degradation of ataxin-2. *Neuroscience letters* 563: 112-7

Ashizawa T, Figueroa KP, Perlman SL, Gomez CM, Wilmot GR, et al. 2013. Clinical characteristics of patients with spinocerebellar ataxias 1, 2, 3 and 6 in the US; a prospective observational study. *Orphanet journal of rare diseases* 8: 177

Auburger G, Diaz GO, Capote RF, Sanchez SG, Perez MP, et al. 1990. Autosomal dominant ataxia: genetic evidence for locus heterogeneity from a Cuban founder-effect population. *American journal of human genetics* 46: 1163-77

Babovic-Vuksanovic D, Snow K, Patterson MC, Michels VV. 1998. Spinocerebellar ataxia type 2 (SCA 2) in an infant with extreme CAG repeat expansion. *American journal of medical genetics* 79: 383-7

Bak RO, Gomez-Ospina N, Porteus MH. 2018. Gene Editing on Center Stage. *Trends Genet* 34: 600-11

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, et al. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315: 1709-12

Bartel DP. 2004. MicroRNAs. *Cell* 116: 281-97

Becker LA, Huang B, Bieri G, Ma R, Knowles DA, et al. 2017. Therapeutic reduction of ataxin-2 extends lifespan and reduces pathology in TDP-43 mice. *Nature* 544: 367-71

Beurdeley M, Bietz F, Li J, Thomas S, Stoddard T, et al. 2013. Compact designer TALENs for efficient genome engineering. *Nat Commun* 4: 1762

Bilen J, Bonini NM. 2007. Genome-wide screen for modifiers of ataxin-3 neurodegeneration in Drosophila. *PLoS genetics* 3: 1950-64

Bobbin ML, Rossi JJ. 2016. RNA Interference (RNAi)-Based Therapeutics: Delivering on the Promise? *Annu Rev Pharmacol Toxicol* 56: 103-22

Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, et al. 2009. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 326: 1509-12

Bolotin A, Quinquis B, Sorokin A, Ehrlich SD. 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* 151: 2551-61

Bonas U, Stall RE, Staskawicz B. 1989. Genetic and structural characterization of the avirulence gene avrBs3 from Xanthomonas campestris pv. vesicatoria. *Mol Gen Genet* 218: 127-36

Brinkman EK, Chen T, Amendola M, van Steensel B. 2014. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Research* 42: e168-e68

Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, et al. 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321: 960-4

Brown AS, Meera P, Altindag B, Chopra R, Perkins EM, et al. 2018. MTSS1/Src family kinase dysregulation underlies multiple inherited ataxias. *Proc Natl Acad Sci U S A* 115: E12407-e16

Buijsen RAM, Toonen LJA, Gardiner SL, van Roon-Mom WMC. 2019. Genetics, Mechanisms, and Therapeutic Progress in Polyglutamine Spinocerebellar Ataxias. *Neurotherapeutics* 16: 263-86

Chai Y, Koppenhafer SL, Bonini NM, Paulson HL. 1999. Analysis of the role of heat shock protein (Hsp) molecular chaperones in polyglutamine disease. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 19: 10338-47

Chang YK, Chen MH, Chiang YH, Chen YF, Ma WH, et al. 2011. Mesenchymal stem cell transplantation ameliorates motor function deterioration of spinocerebellar ataxia by rescuing cerebellar Purkinje cells. *Journal of biomedical science* 18: 54

Chavez A, Scheiman J, Vora S, Pruitt BW, Tuttle M, et al. 2015. Highly efficient Cas9-mediated transcriptional programming. *Nat Methods* 12: 326-8

Chen B, Gilbert LA, Cimini BA, Schnitzbauer J, Zhang W, et al. 2013. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* 155: 1479-91

Chen X, Wu J, Luo Y, Liang X, Supnet C, et al. 2011. Expanded polyglutamine-binding peptoid as a novel therapeutic agent for treatment of Huntington's disease. *Chemistry & biology* 18: 1113-25

Chen ZZ, Wang CM, Lee GC, Hsu HC, Wu TL, et al. 2015. Trehalose attenuates the gait ataxia and gliosis of spinocerebellar ataxia type 17 mice. *Neurochemical research* 40: 800-10

Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, et al. 2010. Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* 186: 757-61

Chuang CY, Yang CC, Soong BW, Yu CY, Chen SH, et al. 2019. Modeling spinocerebellar ataxias 2 and 3 with iPSCs reveals a role for glutamate in disease pathology. *Scientific reports* 9: 1166

Cong L, Ran FA, Cox D, Lin S, Barretto R, et al. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science* 339: 819-23

Cornelius N, Wardman JH, Hargreaves IP, Neergheen V, Bie AS, et al. 2017. Evidence of oxidative stress and mitochondrial dysfunction in spinocerebellar ataxia type 2 (SCA2) patient fibroblasts: Effect of coenzyme Q10 supplementation on these parameters. *Mitochondrion* 34: 103-14

Cummings CJ, Mancini MA, Antalffy B, DeFranco DB, Orr HT, Zoghbi HY. 1998. Chaperone suppression of aggregation and altered subcellular proteasome localization imply protein misfolding in SCA1. *Nat Genet* 19: 148-54

Cummings CJ, Sun Y, Opal P, Antalffy B, Mestril R, et al. 2001. Over-expression of inducible HSP70 chaperone suppresses neuropathology and improves motor function in SCA1 mice. *Hum Mol Genet* 10: 1511-8

Dansithong W, Paul S, Figueroa KP, Rinehart MD, Wiest S, et al. 2015. Ataxin-2 regulates RGS8 translation in a new BAC-SCA2 transgenic mouse model. *PLoS genetics* 11: e1005182

Davis KM, Pattanayak V, Thompson DB, Zuris JA, Liu DR. 2015. Small molecule-triggered Cas9 protein with improved genome-editing specificity. *Nat Chem Biol* 11: 316-8

Di Fabio R, Santorelli F, Bertini E, Balestri M, Cursi L, et al. 2012. Infantile childhood onset of spinocerebellar ataxia type 2. *Cerebellum* 11: 526-30

Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, et al. 2016. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology* 34: 184-91

Doudna JA, Charpentier E. 2014. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* 346: 1258096

Dunbar CE, High KA, Joung JK, Kohn DB, Ozawa K, Sadelain M. 2018. Gene therapy comes of age. *Science* 359

Düzgüneş N, Cheung J, Konopka K. 2018. Non-viral suicide gene therapy in cervical, oral and pharyngeal carcinoma cells with CMV- and EEV-plasmids. *The journal of gene medicine* 20: e3054

Egorova PA, Bezprozvanny IB. 2019. Molecular Mechanisms and Therapeutics for Spinocerebellar Ataxia Type 2. *Neurotherapeutics* 16: 1050-73

Egorova PA, Zakharova OA, Vlasova OL, Bezprozvanny IB. 2016. In vivo analysis of cerebellar Purkinje cell activity in SCA2 transgenic mouse model. *Journal of neurophysiology* 115: 2840-51

Elden AC, Kim HJ, Hart MP, Chen-Plotkin AS, Johnson BS, et al. 2010. Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. *Nature* 466: 1069-75

Estrada R, Galarraga J, Orozco G, Nodarse A, Auburger G. 1999. Spinocerebellar ataxia 2 (SCA2): morphometric analyses in 11 autopsies. *Acta Neuropathol* 97: 306-10

Farg MA, Soo KY, Warraich ST, Sundaramoorthy V, Blair IP, Atkin JD. 2013. Ataxin-2 interacts with FUS and intermediate-length polyglutamine expansions enhance FUS-related pathology in amyotrophic lateral sclerosis. *Hum Mol Genet* 22: 717-28

Fernandez A, Josa S, Montoliu L. 2017. A history of genome editing in mammals. *Mamm Genome* 28: 237-46

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. 1998. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* 391: 806-11

Gafni J, Ellerby LM. 2002. Calpain activation in Huntington's disease. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 22: 4842-9

Gasiunas G, Barrangou R, Horvath P, Siksnys V. 2012. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A* 109: E2579-86

Gatchel JR, Zoghbi HY. 2005. Diseases of unstable repeat expansion: mechanisms and common principles. *Nature reviews. Genetics* 6: 743-55

Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, et al. 2017. Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. *Nature* 551: 464-71

Geschwind DH, Perlman S, Figueroa CP, Treiman LJ, Pulst SM. 1997. The prevalence and wide clinical spectrum of the spinocerebellar ataxia type 2 trinucleotide repeat in patients with autosomal dominant cerebellar ataxia. *American journal of human genetics* 60: 842-50

Gierga K, Burk K, Bauer M, Orozco Diaz G, Auburger G, et al. 2005. Involvement of the cranial nerves and their nuclei in spinocerebellar ataxia type 2 (SCA2). *Acta Neuropathol* 109: 617-31

Gigante AF, Lelli G, Romano R, Pellicciari R, Di Candia A, et al. 2020. The Relationships Between Ataxia and Cognition in Spinocerebellar Ataxia Type 2. *Cerebellum* 19: 40-47

Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, et al. 2013. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 154: 442-51

Giri S, Purushottam M, Viswanath B, Muddashetty RS. 2019. Generation of a FMR1 homozygous knockout human embryonic stem cell line (WAe009-A-16) by CRISPR/Cas9 editing. *Stem cell research* 39: 101494

Goswami R, Subramanian G, Silayeva L, Newkirk I, Doctor D, et al. 2019. Gene Therapy Leaves a Vicious Cycle. *Front Oncol* 9: 297

Guilinger JP, Thompson DB, Liu DR. 2014. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat Biotechnol* 32: 577-82

Guo T, Feng YL, Xiao JJ, Liu Q, Sun XN, et al. 2018. Harnessing accurate non-homologous end joining for efficient precise deletion in CRISPR/Cas9-mediated genome editing. *Genome Biol* 19: 170

Haacke A, Hartl FU, Breuer P. 2007. Calpain inhibition is sufficient to suppress aggregation of polyglutamine-expanded ataxin-3. *The Journal of biological chemistry* 282: 18851-6

Haft DH, Selengut J, Mongodin EF, Nelson KE. 2005. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput Biol* 1: e60

Han H. 2018. RNA Interference to Knock Down Gene Expression. *Methods Mol Biol* 1706: 293-302

Hemphill J, Borchardt EK, Brown K, Asokan A, Deiters A. 2015. Optical Control of CRISPR/Cas9 Gene Editing. *J Am Chem Soc* 137: 5642-5

High KA, Roncarolo MG. 2019. Gene Therapy. *N Engl J Med* 381: 455-64

Hilton IB, D'Ippolito AM, Vockley CM, Thakore PI, Crawford GE, et al. 2015. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol* 33: 510-7

Hsu PD, Lander ES, Zhang F. 2014. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 157: 1262-78

Huynh DP, Figueroa K, Hoang N, Pulst SM. 2000. Nuclear localization or inclusion body formation of ataxin-2 are not necessary for SCA2 pathogenesis in mouse or human. *Nat Genet* 26: 44-50

Huynh DP, Nguyen DT, Pulst-Korenberg JB, Brice A, Pulst SM. 2007. Parkin is an E3 ubiquitin-ligase for normal and mutant ataxin-2 and prevents ataxin-2-induced cell death. *Experimental neurology* 203: 531-41

Huynh DP, Yang H-T, Vakharia H, Nguyen D, Pulst SM. 2003. Expansion of the polyQ repeat in ataxin-2 alters its Golgi localization, disrupts the Golgi complex and causes cell death. *Human Molecular Genetics* 12: 1485-96

Imbert G, Saudou F, Yvert G, Devys D, Trottier Y, et al. 1996. Cloning of the gene for spinocerebellar ataxia 2 reveals a locus with high sensitivity to expanded CAG/glutamine repeats. *Nat Genet* 14: 285-91

Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A. 1987. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J Bacteriol* 169: 5429-33

Jansen R, Embden JD, Gaastra W, Schouls LM. 2002. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* 43: 1565-75

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337: 816-21

Jones J, Jaramillo-Merchan J, Bueno C, Pastor D, Viso-Leon M, Martinez S. 2010. Mesenchymal stem cells rescue Purkinje cells and improve motor functions in a mouse model of cerebellar ataxia. *Neurobiology of disease* 40: 415-23

Ka D, Lee H, Jung YD, Kim K, Seok C, et al. 2016. Crystal Structure of Streptococcus pyogenes Cas1 and Its Interaction with Csn2 in the Type II CRISPR-Cas System. *Structure* 24: 70-79

Kampinga HH, Bergink S. 2016. Heat shock proteins as potential targets for protective strategies in neurodegeneration. *The Lancet. Neurology* 15: 748-59

Kasumu AW, Hougaard C, Rode F, Jacobsen TA, Sabatier JM, et al. 2012. Selective positive modulator of calcium-activated potassium channels exerts beneficial effects in a mouse model of spinocerebellar ataxia type 2. *Chemistry & biology* 19: 1340-53

Kc M, Steer CJ. 2019. A new era of gene editing for the treatment of human diseases. *Swiss medical weekly* 149: w20021

Kearns NA, Pham H, Tabak B, Genga RM, Silverstein NJ, et al. 2015. Functional annotation of native enhancers with a Cas9-histone demethylase fusion. *Nat Methods* 12: 401-03

Keiser MS, Boudreau RL, Davidson BL. 2014. Broad therapeutic benefit after RNAi expression vector delivery to deep cerebellar nuclei: implications for spinocerebellar ataxia type 1 therapy. *Mol Ther* 22: 588-95

Khong A, Matheny T, Jain S, Mitchell SF, Wheeler JR, Parker R. 2017. The Stress Granule Transcriptome Reveals Principles of mRNA Accumulation in Stress Granules. *Molecular cell* 68: 808-20.e5

Kim HJ, Lee HJ, Kim H, Cho SW, Kim JS. 2009. Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. *Genome Res* 19: 1279-88

Kim YG, Cha J, Chandrasegaran S. 1996. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci U S A* 93: 1156-60

Kim YJ, Yi Y, Sapp E, Wang Y, Cuiffo B, et al. 2001. Caspase 3-cleaved N-terminal fragments of wild-type and mutant huntingtin are present in normal and Huntington's disease brains, associate with membranes, and undergo calpain-dependent proteolysis. *Proc Natl Acad Sci U S A* 98: 12784-9

Koonin EV, Makarova KS, Zhang F. 2017. Diversity, classification and evolution of CRISPR-Cas systems. *Current opinion in microbiology* 37: 67-78

Laffita-Mesa JM, Almaguer-Mederos LE, Kouri V, Bauer PO, Vazquez-Mojena Y, et al. 2014. Large normal alleles and SCA2 prevalence: lessons from a nationwide study and analysis of the literature. *Clinical genetics* 86: 96-8

Li PP, Sun X, Xia G, Arbez N, Paul S, et al. 2016. ATXN2-AS, a gene antisense to ATXN2, is associated with spinocerebellar ataxia type 2 and amyotrophic lateral sclerosis. *Annals of neurology* 80: 600-15

Li T, Huang S, Jiang WZ, Wright D, Spalding MH, et al. 2011. TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Res* 39: 359-72

Lin CH, Wu YR, Yang JM, Chen WL, Chao CY, et al. 2016. Novel Lactulose and Melibiose Targeting Autophagy to Reduce PolyQ Aggregation in Cell Models of Spinocerebellar Ataxia 3. *CNS & neurological disorders drug targets* 15: 351-9

Liu J, Tang TS, Tu H, Nelson O, Herndon E, et al. 2009. Deranged calcium signaling and neurodegeneration in spinocerebellar ataxia type 2. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 29: 9148-62

Lorenzetti D, Bohlega S, Zoghbi HY. 1997. The expansion of the CAG repeat in ataxin-2 is a frequent cause of autosomal dominant spinocerebellar ataxia. *Neurology* 49: 1009-13

Maeder ML, Gersbach CA. 2016. Genome-editing Technologies for Gene and Cell Therapy. *Mol Ther* 24: 430-46

Magaña JJ, Velázquez-Pérez L, Cisneros B. 2013. Spinocerebellar ataxia type 2: clinical presentation, molecular mechanisms, and therapeutic perspectives. *Molecular neurobiology* 47: 90-104

Maji B, Moore CL, Zetsche B, Volz SE, Zhang F, et al. 2017. Multidimensional chemical control of CRISPR-Cas9. *Nat Chem Biol* 13: 9-11

Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, et al. 2011. Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* 9: 467-77

Mali P, Esvelt KM, Church GM. 2013a. Cas9 as a versatile tool for engineering biology. *Nat Methods* 10: 957-63

Mali P, Yang L, Esvelt KM, Aach J, Guell M, et al. 2013b. RNA-guided human genome engineering via Cas9. *Science* 339: 823-6

Martí E. 2016. RNA toxicity induced by expanded CAG repeats in Huntington's disease. *Brain pathology (Zurich, Switzerland)* 26: 779-86

Matilla-Duenas A, Sanchez I, Corral-Juan M, Davalos A, Alvarez R, Latorre P. 2010. Cellular and molecular pathways triggering neurodegeneration in the spinocerebellar ataxias. *Cerebellum* 9: 148-66

Matos CA, Carmona V, Vijayakumar UG, Lopes S, Albuquerque P, et al. 2018. Gene Therapies for Polyglutamine Diseases. *Adv Exp Med Biol* 1049: 395-438

Matsumura R, Futamura N. 2001. Late-onset SCA2: 33 CAG repeats are sufficient to cause disease. *Neurology* 57: 566

Meera P, Pulst SM, Otis TS. 2016. Cellular and circuit mechanisms underlying spinocerebellar ataxias. *The Journal of physiology* 594: 4653-60

Miller J, McLachlan AD, Klug A. 1985. Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *The EMBO journal* 4: 1609-14

Miyazaki Y, Du X, Muramatsu S, Gomez CM. 2016. An miRNA-mediated therapy for SCA6 blocks IRES-driven translation of the CACNA1A second cistron. *Science translational medicine* 8: 347ra94

Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Soria E. 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* 60: 174-82

Mojica FJ, Juez G, Rodriguez-Valera F. 1993. Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites. *Mol Microbiol* 9: 613-21

Moscou MJ, Bogdanove AJ. 2009. A simple cipher governs DNA recognition by TAL effectors. *Science* 326: 1501

Nagai Y, Fujikake N, Ohno K, Higashiyama H, Popiel HA, et al. 2003. Prevention of polyglutamine oligomerization and neurodegeneration by the peptide inhibitor QBP1 in Drosophila. *Hum Mol Genet* 12: 1253-9

Nascimento-Ferreira I, Santos-Ferreira T, Sousa-Ferreira L, Auregan G, Onofre I, et al. 2011. Overexpression of the autophagic beclin-1 protein clears mutant ataxin-3 and alleviates Machado-Joseph disease. *Brain : a journal of neurology* 134: 1400-15

Nihei Y, Ito D, Suzuki N. 2012. Roles of ataxin-2 in pathological cascades mediated by TAR DNA-binding protein 43 (TDP-43) and Fused in Sarcoma (FUS). *The Journal of biological chemistry* 287: 41310-23

Nihongaki Y, Kawano F, Nakajima T, Sato M. 2015. Photoactivatable CRISPR-Cas9 for optogenetic genome editing. *Nat Biotechnol* 33: 755-60

Nishida K, Arazoe T, Yachie N, Banno S, Kakimoto M, et al. 2016. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* 353

Nishiyama J, Mikuni T, Yasuda R. 2017. Virus-Mediated Genome Editing via Homology-Directed Repair in Mitotic and Postmitotic Cells in Mammalian Brain. *Neuron* 96: 755-68.e5

Nóbrega C, Nascimento-Ferreira I, Onofre I, Albuquerque D, Hirai H, et al. 2013. Silencing mutant ataxin-3 rescues motor deficits and neuropathology in Machado-Joseph disease transgenic mice. *PLoS One* 8: e52396

Nóbrega C, Pereira de Almeida L. 2018. *Polyglutamine Disorders*. Springer International Publishing. VIII, 469 pp.

Nonhoff U, Ralser M, Welzel F, Piccini I, Balzereit D, et al. 2007. Ataxin-2 interacts with the DEAD/H-box RNA helicase DDX6 and interferes with P-bodies and stress granules. *Molecular biology of the cell* 18: 1385-96

Nonis D, Schmidt MHH, van de Loo S, Eich F, Dikic I, et al. 2008. Ataxin-2 associates with the endocytosis complex and affects EGF receptor trafficking. *Cellular signalling* 20: 1725-39

Nunez JK, Kranzusch PJ, Noeske J, Wright AV, Davies CW, Doudna JA. 2014. Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nat Struct Mol Biol* 21: 528-34

Ouyang S, Xie Y, Xiong Z, Yang Y, Xian Y, et al. 2018. CRISPR/Cas9-Targeted Deletion of Polyglutamine in Spinocerebellar Ataxia Type 3-Derived Induced Pluripotent Stem Cells. *Stem cells and development* 27: 756-70

Paul S, Dansithong W, Figueroa KP, Scoles DR, Pulst SM. 2018. Staufen1 links RNA stress granules and autophagy in a model of neurodegeneration. *Nat Commun* 9: 3648

Platt RJ, Chen S, Zhou Y, Yim MJ, Swiech L, et al. 2014. CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell* 159: 440-55

Pulst SM, Nechiporuk A, Nechiporuk T, Gispert S, Chen XN, et al. 1996. Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. *Nat Genet* 14: 269-76

Puorro G, Marsili A, Sapone F, Pane C, De Rosa A, et al. 2018. Peripheral markers of autophagy in polyglutamine diseases. *Neurological sciences : official journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology* 39: 149-52

Ralser M, Albrecht M, Nonhoff U, Lengauer T, Lehrach H, Krobitsch S. 2005a. An integrative approach to gain insights into the cellular function of human ataxin-2. *J Mol Biol* 346: 203-14

Ralser M, Nonhoff U, Albrecht M, Lengauer T, Wanker EE, et al. 2005b. Ataxin-2 and huntingtin interact with endophilin-A complexes to function in plastin-associated pathways. *Hum Mol Genet* 14: 2893-909

Ramocki MB, Chapieski L, McDonald RO, Fernandez F, Malphrus AD. 2008. Spinocerebellar ataxia type 2 presenting with cognitive regression in childhood. *Journal of child neurology* 23: 999-1001

Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, et al. 2015. In vivo genome editing using Staphylococcus aureus Cas9. *Nature* 520: 186-91

Rath D, Amlinger L, Rath A, Lundgren M. 2015. The CRISPR-Cas immune system: biology, mechanisms and applications. *Biochimie* 117: 119-28

Rebar EJ, Pabo CO. 1994. Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. *Science* 263: 671-3

Reis SD, Pinho BR, Oliveira JMA. 2017. Modulation of Molecular Chaperones in Huntington's Disease and Other Polyglutamine Disorders. *Molecular neurobiology* 54: 5829-54

Ross OA, Rutherford NJ, Baker M, Soto-Ortolaza AI, Carrasquillo MM, et al. 2011. Ataxin-2 repeat-length variation and neurodegeneration. *Human Molecular Genetics* 20: 3207-12

Rub U, Seidel K, Ozerden I, Gierga K, Brunt ER, et al. 2007. Consistent affection of the central somatosensory system in spinocerebellar ataxia type 2 and type 3 and its significance for clinical symptoms and rehabilitative therapy. *Brain Res Rev* 53: 235-49

Sanjana NE, Shalem O, Zhang F. 2014. Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods* 11: 783-84

Sanpei K, Takano H, Igarashi S, Sato T, Oyake M, et al. 1996. Identification of the spinocerebellar ataxia type 2 gene using a direct identification of repeat expansion and cloning technique, DIRECT. *Nat Genet* 14: 277-84

Satterfield TF, Pallanck LJ. 2006. Ataxin-2 and its Drosophila homolog, ATX2, physically assemble with polyribosomes. *Hum Mol Genet* 15: 2523-32

Saudou F, Finkbeiner S, Devys D, Greenberg ME. 1998. Huntingtin acts in the nucleus to induce apoptosis but death does not correlate with the formation of intranuclear inclusions. *Cell* 95: 55-66

Scoles DR, Meera P, Schneider MD, Paul S, Dansithong W, et al. 2017. Antisense oligonucleotide therapy for spinocerebellar ataxia type 2. *Nature* 544: 362-66

Scoles DR, Pulst SM. 2018. Spinocerebellar Ataxia Type 2. *Adv Exp Med Biol* 1049: 175-95

Segal DJ, Beerli RR, Blancafort P, Dreier B, Effertz K, et al. 2003. Evaluation of a modular strategy for the construction of novel polydactyl zinc finger DNA-binding proteins. *Biochemistry* 42: 2137-48

Seidel K, Siswanto S, Brunt ER, den Dunnen W, Korf HW, Rub U. 2012. Brain pathology of spinocerebellar ataxias. *Acta Neuropathol* 124: 1-21

Seligman LM, Chisholm KM, Chevalier BS, Chadsey MS, Edwards ST, et al. 2002. Mutations altering the cleavage specificity of a homing endonuclease. *Nucleic Acids Res* 30: 3870-9

Sen NE, Drost J, Gispert S, Torres-Odio S, Damrath E, et al. 2016. Search for SCA2 blood RNA biomarkers highlights Ataxin-2 as strong modifier of the mitochondrial factor PINK1 levels. *Neurobiology of disease* 96: 115-26

Setten RL, Rossi JJ, Han SP. 2019. The current state and future directions of RNAi-based therapeutics. *Nature reviews. Drug discovery* 18: 421-46

Shibata H, Huynh DP, Pulst SM. 2000. A novel protein with RNA-binding motifs interacts with ataxin-2. *Hum Mol Genet* 9: 1303-13

Shin JW, Kim KH, Chao MJ, Atwal RS, Gillis T, et al. 2016. Permanent inactivation of Huntington's disease mutation by personalized allele-specific CRISPR/Cas9. *Hum Mol Genet* 25: 4566-76

Silva G, Poirot L, Galetto R, Smith J, Montoya G, et al. 2011. Meganucleases and other tools for targeted genome engineering: perspectives and challenges for gene therapy. *Curr Gene Ther* 11: 11-27

Simoes AT, Goncalves N, Koeppen A, Deglon N, Kugler S, et al. 2012. Calpastatin-mediated inhibition of calpains in the mouse brain prevents mutant ataxin 3 proteolysis, nuclear localization and aggregation, relieving Machado-Joseph disease. *Brain : a journal of neurology* 135: 2428-39

Soubeyran P, Kowanetz K, Szymkiewicz I, Langdon WY, Dikic I. 2002. Cbl-CIN85-endophilin complex mediates ligand-induced downregulation of EGF receptors. *Nature* 416: 183-7

Stepper P, Kungulovski G, Jurkowska RZ, Chandra T, Krueger F, et al. 2017. Efficient targeted DNA methylation with chimeric dCas9-Dnmt3a-Dnmt3L methyltransferase. *Nucleic Acids Res* 45: 1703-13

Stolberg SG. 1999. The biotech death of Jesse Gelsinger. *The New York times magazine*: 136-40, 49-50

Sugisaki H, Kanazawa S. 1981. New restriction endonucleases from Flavobacterium okeanokoites (FokI) and Micrococcus luteus (MluI). *Gene* 16: 73-8

Sum CH, Shortall SM, Wong S, Wettig SD. 2018. Non-viral Gene Delivery. *Exp Suppl* 110: 3-68

Tan JY, Vance KW, Varela MA, Sirey T, Watson LM, et al. 2014. Cross-talking noncoding RNAs contribute to cell-specific neurodegeneration in SCA7. *Nat Struct Mol Biol* 21: 955-61

Thakore PI, D'Ippolito AM, Song L, Safi A, Shivakumar NK, et al. 2015. Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat Methods* 12: 1143-9

Tsai YA, Liu RS, Lirng JF, Yang BH, Chang CH, et al. 2017. Treatment of Spinocerebellar Ataxia With Mesenchymal Stem Cells: A Phase I/IIa Clinical Study. *Cell transplantation* 26: 503-12

Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD. 2010. Genome editing with engineered zinc finger nucleases. *Nature Reviews Genetics* 11: 636-46

Velázquez-Pérez LC, Rodríguez-Labrada R, Fernandez-Ruiz J. 2017. Spinocerebellar Ataxia Type 2: Clinicogenetic Aspects, Mechanistic Insights, and Management Approaches. *Frontiers in neurology* 8: 472

Velazquez Perez L, Cruz GS, Santos Falcon N, Enrique Almaguer Mederos L, Escalona Batallan K, et al. 2009. Molecular epidemiology of spinocerebellar ataxias in Cuba: insights into SCA2 founder effect in Holguin. *Neuroscience letters* 454: 157-60

Verbeek DS, Goedhart J, Bruinsma L, Sinke RJ, Reits EA. 2008. PKC gamma mutations in spinocerebellar ataxia type 14 affect C1 domain accessibility and kinase activity leading to aberrant MAPK signaling. *Journal of cell science* 121: 2339-49

Wadia NH, Swami RK. 1971. A new form of heredo-familial spinocerebellar degeneration with slow eye movements (nine families). *Brain : a journal of neurology* 94: 359-74

Wan L, Xu K, Chen Z, Tang B, Jiang H. 2018. Roles of Post-translational Modifications in Spinocerebellar Ataxias. *Frontiers in cellular neuroscience* 12: 290

Wang LC, Chen KY, Pan H, Wu CC, Chen PH, et al. 2011. Muscleblind participates in RNA toxicity of expanded CAG and CUG repeats in Caenorhabditis elegans. *Cellular and molecular life sciences : CMLS* 68: 1255-67

Warrick JM, Chan HY, Gray-Board GL, Chai Y, Paulson HL, Bonini NM. 1999. Suppression of polyglutamine-mediated neurodegeneration in Drosophila by the molecular chaperone HSP70. *Nat Genet* 23: 425-8

Williams AJ, Paulson HL. 2008. Polyglutamine neurodegeneration: protein misfolding revisited. *Trends in neurosciences* 31: 521-8

Wirth T, Parker N, Yla-Herttuala S. 2013. History of gene therapy. *Gene* 525: 162-9

Xu L, Wang J, Liu Y, Xie L, Su B, et al. 2019. CRISPR-Edited Stem Cells in a Patient with HIV and Acute Lymphocytic Leukemia. *N Engl J Med* 381: 1240-47

Yu Z, Zhu Y, Chen-Plotkin AS, Clay-Falcone D, McCluskey L, et al. 2011. PolyQ repeat expansions in ATXN2 associated with ALS are CAA interrupted repeats. *PLoS One* 6: e17951

Zetsche B, Volz SE, Zhang F. 2015. A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nat Biotechnol* 33: 139-42

Zhang XH, Tee LY, Wang XG, Huang QS, Yang SH. 2015. Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. *Mol Ther Nucleic Acids* 4: e264