Bernardo Alexandre Encarnação Estevam

PolyQ Database – A polyglutamine disease database

UNIVERSIDADE DO ALGARVE

Faculdade de Medicina e Ciências Biomédicas

2022

Bernardo Alexandre Encarnação Estevam

PolyQ Database – A polyglutamine disease database

UNIVERSIDADE DO ALGARVE

Faculdade de Medicina e Ciências Biomédicas

2022

PolyQ Database – A polyglutamine disease database

Declaração de autoria de trabalho

Declaro ser o autor deste trabalho, que é original e inédito. Autores e trabalhos consultados estão devidamente citados no texto e constam na listagem de referências incluída.

Authorship statement

I hereby declare to be the author of this work, which is original and unpublished.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

_____

(Bernardo Estevam)

AGRADECIMENTOS

A realização deste projeto foi extremamente interessante e marcante no aspeto em que finaliza um passo académico importante. Importante frisar que sem o apoio de várias pessoas, não teriam sido cumpridos todos os objetivos. A essas pessoas deixo aqui os meus sinceros agradecimentos.

Ao meu orientador, Professor Doutor Clévio Nóbrega, por me ter apresentado este projeto e ter sido o meu mentor durante ao longo do mesmo. Pela disponibilidade e pela excelência científica que sempre me mostrou assim como a paciência na resolução de pontuais problemas.

Ao meu coorientador, Professor Carlos Matos, pela excelente ajuda científica graças ao seu *background* profissional e pela sugestão de ferramentas que facilitaram a resolução de vários problemas. Pelas críticas construtivas ao longo do projeto e também pela disponibilidade e paciência.

Um agradecimento à Dra. Isa Mestre pela ajuda no lançamento do website, sendo intermediária entre mim e as empresas fornecedoras do pretendido serviço.

Especial agradecimento ao Patrick Arsénio pela excelente ajuda no *backend* e na configuração do website para produção que se mostrou um difícil desafio devido a eu não possuir qualquer *background* na área. Pela disponibilidade e pela paciência para me ensinar tudo o que era suposto eu saber. Pelas constantes recomendações e diversas resoluções de problemas que iam aparecendo. Posso dizer que sem a ajuda do Patrick provavelmente este site ainda não estava feito daí receber o meu sincero obrigado.

Sem esquecer um agradecimento à minha família e amigos pelo apoio e por estarem sempre presentes tanto nesta etapa académica como noutros momentos.

TABLE OF CONTENTS

INDEX OF FIGURES

INDEX OF TABLES

LIST OF ABBREVIATIONS

ADCA - Autosomal dominant cerebellar ataxias

ALS – Amyotrophic lateral sclerosis

AR – Androgen receptor

Cic - Capicua

CREB – cAMP response element-binding protein

CSS – Cascading style sheets

DBD – DNA binding domains

DBMS – Database management system

DRPLA – Dentatorubral-pallidoluysian atrophy

EGFR – Epidermal growth factor receptor

FAT - Fast axonal support

FRDA – Friedreich Ataxia

GCN5 - General control non-depressible 5

HD – Huntington's disease

HTML – Hypertext markup language

HTT – Huntingtin

JS – JavaScript

KAT2A - Lysine acetyltransferase 2A

LC3 - Microtubule-associated protein 1 light chain 3

LSD1 - Lysine-specific demethylase 1

Mbnl1 - muscleblind-link-1

MJD – Machado Joseph disease

NCBI - National center for biotechnology information

NES – Nuclear export signal

NI – Nuclear inclusion

NLS – Nuclear localization signal

NMDA - N-methyl-D-aspartate

NNI – Neuronal nuclear inclusion

OPCA - Olivopontocerebellar atrophy

PIC – Pre-initiation complex

PolyQ – Polyglutamine

SAGA - Spt-Ada-Gcn5 acetyltransferase

SBMA - Spinal and bulbar muscular atrophy

SCA – Spinocerebellar ataxia

SOD – Superoxide dismutase

Sp1 – Tpecific-protein 1

SSM – Slipped strand mispairing

STAU1 - Staufen1

TAFII130 - Transcription initiation factor TFIID subunit 4

TBP – TATA binding protein

TF – Transcription factor

TFIID - Transcription factor II D

TNR – Trinucleotide repeat expansion

UPS – Ubiquitin proteasome system

VGCC - Voltage-gated Ca2+ channels

XML - Extensible Markup Language

ABSTRACT

Polyglutamine diseases are neurodegenerative disorders where the associated genes and translated protein products have an abnormal number of CAG triplets and glutamines, respectively. Mutated proteins with increased number of glutamines are usually correlated with neuronal dysfunction. Depending on the length of the pathogenic mutation, the effects on the normal cell mechanisms may be less or more severe, in other words, the severity of the disease is usually directly proportional with the number of glutamine repeats.

There are 9 polyglutamine diseases: Huntington's disease (HD), dentatorubral-pallidoluysian atrophy (DRPLA), spinal and bulbar muscular atrophy (SBMA) and the spinocerebellar ataxias (SCAs) 1, 2, 3, 6, 7 and 17. All these diseases are caused by the same mutation in their respective, and otherwise unrelated, genes – CAG triplet expansion. This abnormal triplet expansion is found in coding gene regions and later gives rise to the expanded glutamine sequence in the translated proteins. Mutated proteins with an increased number of glutamines often interfere with the normal cellular function by not performing their normal tasks and by compromising crucial cell mechanisms.

PolyQ Database is a project that aims to be an online resource where everyone can learn about the most important features of every polyglutamine disease. Since information on each specific disease is usually scattered throughout various online sources, one of the main goals is to have all the most important information in one single online resource. The data available in PolyQ Database consists of simple and straightforward introductory information followed by more complex explanations about the genes and proteins affected in each disease, along with the pathophysiological mechanisms and cellular and biological deficits that may arise from these.

This platform was done with hypertext markup language, cascading style sheets and JavaScript for the front-end (what the user interacts with). The back-end structure was entirely made with python Django framework. All the data stored for each disease was initially stored in an object oriented SQlite3 database created with Django but was later imported to a Microsoft SQL server.

To obtain a scientific database structure with consistent and constant slices of information, the same data was presented for each polyglutamine disease. For each condition, the topics presented are: introduction, first description, epidemiology, causative gene, codified protein (structure, domains, functions and intracellular localization), pathophysiology and clinical manifestations (neuropathology, other signs and symptoms).

The PolyQ Database can be found at https://polyq.pt/, where all the information gathered from various sources is presented, along with added supplementary pages, including a contact and authors pages, and a home page to search the diseases using either a search engine or disease tags.

Keywords: Huntington's disease (HD), dentatorubral-pallidoluysian atrophy (DRPLA), spinal and bulbar muscular atrophy (SBMA), spinocerebellar ataxia (SCA)

RESUMO

As doenças de poliglutaminas são um conjunto de doenças neurodegenerativas associadas à presença de sequências anormalmente longas de glutaminas em proteínas particulares. Este número elevado de glutaminas é consequência da repetição anormal de tripletos CAG existentes na parte codificante de genes respetivos, que são posteriormente traduzidos para proteína. Sequências de repetições de CAG com maior tamanho estão diretamente relacionados com uma maior probabilidade de desenvolver doença e com uma severidade das manifestações clínicas. O mecanismo principal que parece estar na origem destas doenças é a expansão de repetições do trinucleotídeo CAG, em que expansões anormais surgem de mutações indel, ou simplesmente, mutações de inserção ou deleção. O principal mecanismo por detrás destas mutações indel é o *slipped strand mispairing* (SSM). SSM é um tipo de mutação que pode ocorrer durante a replicação de DNA e que envolve vários processos que causam mutações ligeiramente diferentes. No caso das doenças de poliglutaminas, a repetição anormal de CAG dever-se-á ao facto de a DNA polimerase, ao sintetizar a nova cadeia, dar "um salto" para trás e copiar novamente o que já foi copiado.

As proteínas traduzidas a partir destes genes mutados podem comprometer vários processos celulares, que estão relacionados com o desenvolvimento das respetivas doenças. Proteínas com um número elevado de glutaminas são geralmente menos eficientes a executar as suas funções normais comprometendo assim vários processos celulares em que participam. Com o aumento anormal de glutaminas, existe também a possibilidade de as proteínas se envolverem em novos processos celulares, com os quais, em situações normais, não se envolveriam. Outra característica de diversas doenças de poliglutaminas é a existência de agregados citoplasmáticos e nucleares (inclusões nucleares) que contêm a proteína mutada. Estas agregados podem sequestrar outras proteínas importantes e criar agregados macromoleculares que a célula não tem capacidade de degradar. A formação de agregados/inclusões no interior do núcleo pode interferir com o normal funcionamento de vários processos nucleares, nomeadamente a transcrição.

Existem nove doenças de poliglutaminas extensamente relatadas e identificadas: doença de Huntington, atrofia dentatorubro-palidoluisiana, atrofia muscular espinhal-bulbar e as ataxias espinocerebelosas 1, 2, 3, 6, 7 e 17. Cada uma destas doenças é causada pela

expansão de um gene e de uma proteína particulares, e está associada a degeneração de zonas específicas do sistema nervoso e a mecanismos patofisiológicos distintos, apresentando por isso diferentes sinais pelos quais é possível diferenciar e classificar cada uma das doenças de poliglutaminas. No entanto, considera-se que certos mecanismos patogénicos e sinais neuropatológicos são comuns às nove doenças.

As manifestações clínicas das doenças de poliglutaminas são diversas, e variam de acordo com as estruturas afetadas. Por exemplo, a coreia, no caso da doença de Huntington, e a ataxia, no caso das ataxias espinocerebelosas, surgem respetivamente da decadência funcional dos gânglios basais e do cerebelo, respetivamente. No entanto, estas doenças acabam também por exibir sintomas em comum, que normalmente surgem no decorrer da doença ou em fases mais tardias, como a demência, a fraqueza muscular, epilepsia e convulsões e deficiências cognitivas.

O projeto *PolyQ Database* teve como objetivo o estabelecimento de uma plataforma que permitisse um acesso fácil a todas as informações importantes acerca de cada uma das doenças de poliglutaminas, num mesmo local.

Essa plataforma foi feita com *hypertext markup language*, *cascading style sheets* e JavaScript para o *front-end* (com o que o usuário interage). A estrutura de *back-end* foi inteiramente feita com o *framework* Django do Python. Todos os dados armazenados para cada doença foram inicialmente armazenados num banco de dados SQlite3 orientado a objetos criado com Django, mas posteriormente foram importados para um servidor Microsoft SQL.

De acordo com a estrutura estabelecida, foi desenhado um website que apresenta, para cada uma das 9 doenças, uma parte introdutória seguida pela informação acerca das suas primeiras descrições históricas e por dados epidemiológicos. De seguida, são apresentadas diversas informações acerca do gene afetado e da proteína traduzida (estrutura, funções e localização intracelular). Por fim, são apresentados os mecanismos celulares afetados que levam ao desenvolvimento da doença e os sinais/sintomas observados em pessoas afetadas.

A *PolyQ Database* está disponível em https://polyq.pt/ ,onde se apresenta toda a informação recolhida de várias fontes, de acordo com a estrutura descrita. O website apresenta ainda outras páginas suplementares, como a página de contato e página dos autores. A página *home* permite ao utilizador aceder à informação acerca das diversas

doenças através de um motor de busca ou por escolha direta de qualquer doença disponível por marcadores opcionais.

Palavras chave: *slipped strand mispairing* **(SSM)**

CHAPTER 1 – INTRODUCTION

1. INTRODUCTION

1.1 What are polyglutamine diseases? General information

Polyglutamine (polyQ) diseases are a group of neurodegenerative diseases. As the name implies, these often called polyglutamine disorders are associated with the presence of an abnormally expanded number of amino acid glutamine residues in the particular proteins that leads to neuronal function impairments. Glutamine-expanded proteins are translated from mutant forms of genes that have an amount of cytosine-adenine-guanine (CAG) repeats beyond what is normally found in the healthy population. The range of CAG repeats varies between diseases, in some cases, initial disease manifestations can start when the repeats go past 19, in other cases, when it crosses 48. Since the expanded repeats are found in the coding regions of the genes, they are carried to the RNA transcripts and give rise to mutated proteins. PolyQ-expanded proteins are believed to have a tendency to misfold as a result of their long stretch of glutamines and they self-aggregate, hindering their cellular degradation (figure.1).

Diseases in this group are genetically inherited and more severe and earlier phenotypes are usually associated with longer of repeats expansions (Gatchel & Zoghbi, 2005; H. T. Orr, 2001). Nine polyglutamine diseases have been described to this date: Huntington's disease (HD), dentatorubral-pallidoluysian atrophy (DRPLA), spinal and bulbar muscular atrophy (SBMA) and the spinocerebellar ataxias (SCAs) 1, 2, 3, 6, 7 and 17 (T. Takahashi et al., 2010). Diseases manifestations observed in each disease reflect the functional compromise of the particular regions of the nervous system that undergo degeneration in each particular polyQ disorder. For example, all 6 polyQ SCAs present themselves with ataxia as a consequence of cerebellar degeneration, while chorea observed in HD predominantly results from basal ganglia function decline (Reiner et al., 2011; T. Takahashi et al., 2010). Symptoms are usually within the category of movement impairments, but beyond chorea and ataxia, cognitive impairments and muscular atrophies are also present in some polyQ diseases. Late-onset symptoms of most polyglutamine disorders include dementia and other psychiatric problems (Todd & Lim, 2013). Table 1 groups information on every disease with data obtained from Shao & Diamond, 2007; Takahashi et al., 2010; Tunalı, 2020.

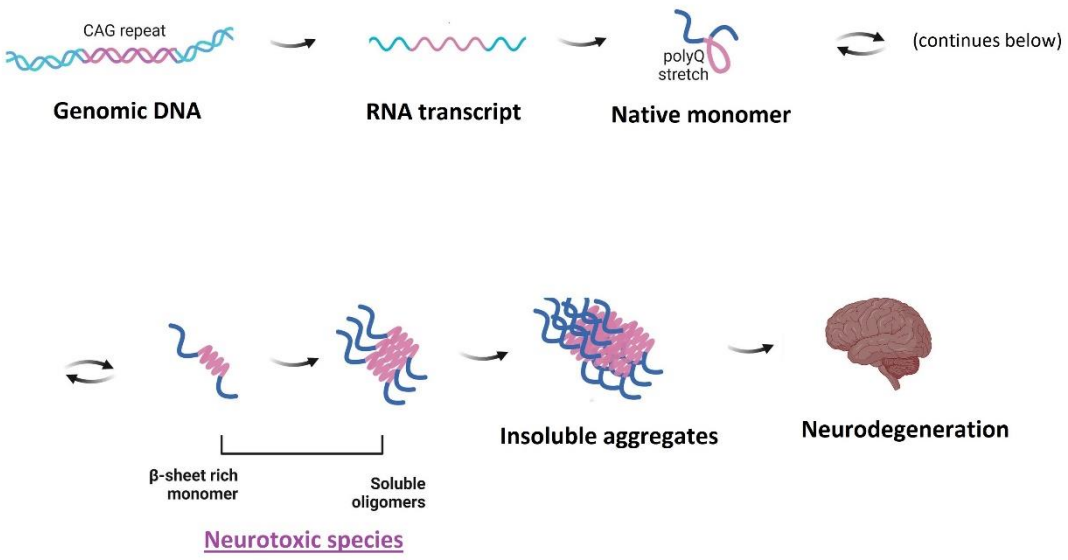

Figure 1: PolyQ disease step by step pathogensis until neurodegeneration. A mutation on certain genes increases number of CAG triplets. When the gene is transcripted, RNA contains expanded repeats. Translation leads to a protein with a long stretch of glutamines. Mutated forms of the protein cannot perform its normal functions as they normally would, interfere with diverse cell cystems and are harder to be destroyed by the cell. Overtime, polyQ proteins aggregate into toxic soluble oligomers and later into insoluble macromolecular deposits. Agreggation along with other intracellular impairments lead to disease progression (adapted from Minakawa & Nagai, 2021).

Neuropathological findings in every polyQ disease are very similar. Because of this similarity many bibliographic sources and authors interpret these findings as polyQ diseases hallmarks. A predominant hallmark is the presence of inclusion bodies in the nucleus and/or cytoplasm of neurons (Takeuchi & Nagai, 2017) (Todd & Lim, 2013).

1.1.1 Trinucleotide repeat expansion.

Trinucleotide repeat (TNR) expansions or triplet expansions (Hegde & Saraph, 2011; La Spada & Taylor, 2010; McMurray, 2010) are mutations in which the number of triplets present in a mutated gene is greater than the number found in a normal gene and this number may continue to increase after being inherited (Budworth & McMurray, 2013a). Four types of TNRs have been identified and documented in diseases: cytosine-guanine-guanine (CGG) in fragile X syndromes, cytosine-thymine-guanine (CTG) in myotonic

Table 1. Characteristics of the 9 polyglutamine diseases including gene, protein, number of CAG repeats in disease, affected tissues/areas and most predominant clinical symptoms.

Disease	Gene/Protein	(CAG)n in disease	Affected areas	Main clinical symptoms
HD	*HTT*/Huntingtin	> 36	Basal ganglia (striatum), cerebral cortex	Chorea
DRPLA	*ATN1*/Atrophin-1	≥ 48	Dentate nucleus, basal ganglia	Ataxia, choreoathetosis, myoclonus, seizures and epilepsy
SBMA	*AR*/Androgen-receptor	> 47	Skeletal muscle, motor neurons of the anterior horn and bulbar regions	Bulbar muscle weakness, atrophy and fasticulation, hormonal dysfunction
SCA1	*ATXN1*/Ataxin-1	> 39	Cerebellum (Purkinje cells, dentate nucleus), pons, pyramidal tracts	Ataxia, dysarthia, progressive motor deterioration
SCA2	*ATXN2*/Ataxin-2	> 32	Cerebellum, pons, medulla oblongata, basal ganglia, frontal lobe, substantia nigra	Ataxia decreased reflexes, dysarthia, parkinsonian rigidity
SCA3	*ATXN3*/Ataxin-3	≥ 53	Cerebellum (dentate nucleus), pons, striatum, substantia nigra	Ataxia, dystonia. parkinsonism, dysarthria, dysphagia, loss of muscle
SCA6	*CACNA1A*/α1A subunit of the voltage-gated Ca2+ channel	> 19	Cerebellum (Purkinje cells)	Ataxia, dysarthria, oculomotor disorders, dysphagia, choking
SCA7	*ATXN7*/Ataxin-7	> 36	Cerebellum, inferior olive, visual cortex	Ataxia, retinal degeneration, dysphagia, dysarthria,
SCA17	*TBP*/TATA-box binding protein	≥ 45	Cerebellum (Purkinje cells)	Ataxia, dementia, epilepsy, seizures

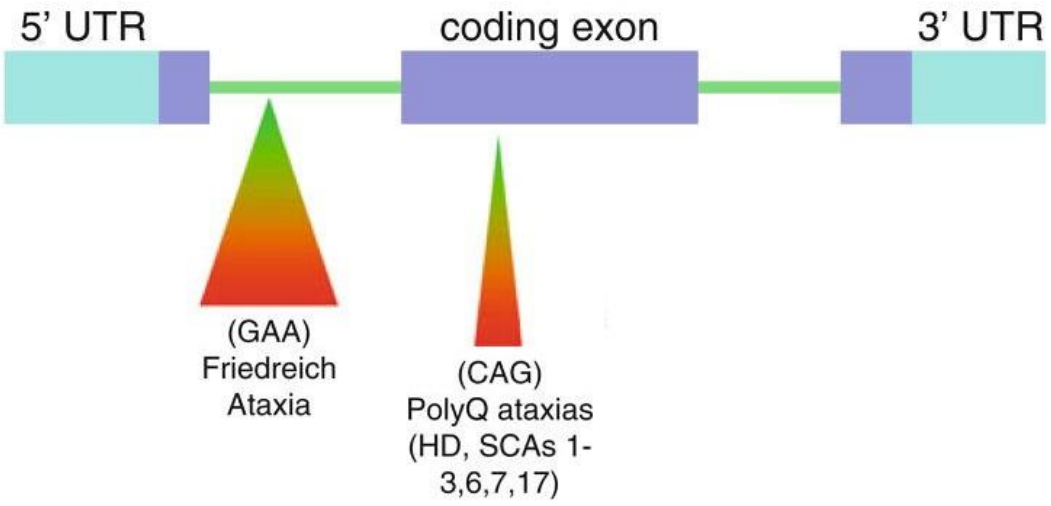

Figure 2: TNRs mutations found in either coding exons or noncoding regions. Most CAG expanded affected diseases have TNRs in their coding exons leading to conformational changes in the resulting protein. Friedreich ataxias presents itself with TNRs in noncoding regions which alters protein expression, in some cases, even suppress it (adapted from Budworth & McMurray, 2013b).

dystrophy, guanine-adenine-adenine in Friedreich's ataxia (GAA) and CAG in polyQ diseases and other disorders (Koshy & Zoghbi, 1997).

In the case of polyQ diseases, the abnormal expansion of the trinucleotide CAG is found in the coding region of the affected genes. As mentioned previously, CAG TNRs residing in a coding sequence of a gene translate into a defective protein (Budworth & McMurray, 2013b). In all polyQ diseases, the TNRs are found within coding exons and are later

translated as the expanded glutamine stretches found on mutated proteins. On the other hand, in some other TNR derived diseases, the expansions are found in non-coding exons/regions, the mutant protein expression may be suppressed, its splicing altered or even have its antisense regulation dysregulated (Hegde & Saraph, 2011; La Spada & Taylor, 2010)(figure 2).

1.1.2 How do trinucleotide repeat expansions occur?

Trinucleotide repeat disorders, in which polyQ diseases are included, are a result of DNA mutations occasioning triplet expansions. These segments of repeated DNA nucleotides are denominated microsatellites. According to A. de Koning *et al*, approximately two

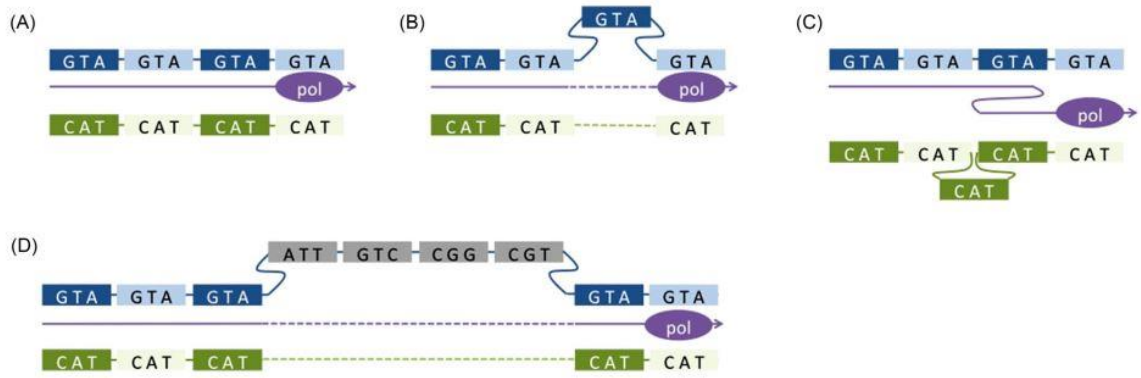

Figure 3: Slipped strand mispairing mechanism for indel mutations of tandem repeats. (A) Normal replication by polymerase. (B) Polymerase skips forward when it reassembles to template leading to deletion of a repeat in the newly generated DNA strand. (C) Polymerase slips backward leading to the insertion of a new extra repeat that was previously synthesized (2 CAT repeats instead of 1). (D) Polymerase can either skip or slip backward repeats not adjacent to the spot it was on before reassembling, leading to deletion or addition of larger numbers of repeats (image from Sehn, 2015).

thirds of the human genome consists of repetitive DNA (de Koning et al., 2011). Repetitive DNA varies in length but microsatellites are possibly the simplest to understand (Kinney et al., 2019). Consisting of short motifs (1-6 base pairs), microsatellites are nowadays contemplated as genetic fingerprints because of their high mutation rates often seen across a variety of diseases, such as SCAs and HD (Kinney et al., 2019; Murmann et al., 2018).

DNA indel mutations are the cause of either expanded or shortened triplet repeats. The most common mechanism for indel occurrence is called slipped strand mispairing (SSM), also known as replication slippage or polymerase slippage (Casorelli et al., 2012; Sehn, 2015) and it mostly affects DNA regions with direct repeats (tandem repeats). During the replication process, the DNA polymerase and the new DNA strand complex sometimes dissociates from the template DNA for short periods and when they reassemble the replication can start positions ahead or before from where it left off leading to insertion and deletion of repeats, respectively (Sehn, 2015)(figure 3).

In the case of polyQ disorders, CAG tracts of the genes are expanded because of SSM backward slippages.

Although the most common explanation for the appearance of repeated triplets resides on SSM, aberrant DNA repair mechanisms are also interpreted as important candidates for creation of expanded trinucleotides in the DNA. It is believed that faulty DNA repair plays an important role in the generation of trinucleotide expansions, this is backed by the fact neurons, being the mostly affected cell type across these disorders, are post-mitotic cells, therefore it has been suggested that expansions found within neuronal cells may be independent of chromosomal replication (van den Broek et al., 2007; X. N. Zhao & Usdin, 2015).

1.1.3 DNA with expanded CAG presents hairpins

In all polyQ disorders, the expanded CAG repeats are located within coding regions of the gene (exons), leading to conformational changes and other alterations on the translated protein (Koshy & Zoghbi, 1997).

Triplet expansions have also been shown to facilitate the formation of certain structures in the DNA. The expanded TNR CAG microsatellite sequences form imperfectly base paired hairpins that expand *in vivo* depending on their length (Liu & Leffak, 2012).

As presented previously, TNR expansions may arise either during DNA replication or repair. Polymerase slippage or SSM is a likely suspect in the formation of hairpins in mitotic cells (Mirkin, 2007; Pearson et al., 2005). However, hairpins may also be generated from other cellular mechanisms such as replication fork reversal and post replication repair (Kerrest et al., 2009; I. V. Kovtun et al., 2004), along with other repair mechanisms which include base excision repair (Irina V. Kovtun et al., 2007), nucleotide excision repair (Lahue & Slater, 2003; Lin et al., 2009) and structure repair induced by R-loop formation during transcriptions (Lin et al., 2009; Nakamori et al., 2011). Models of (CTG)n•(CAG)n instability during replication or repair predict that TNR expansion occurs if newly synthesized DNA strands have a hairpin that is long enough to serve as model for replication and stable hairpin formation in template DNA strands lead to contraction of the repeat on the second round of replication (Liu et al., 2010; Liu & Leffak, 2012)(figure 4). Some forms of these hairpins have been implicated in aggregation and oligomerization because of their stability and dimerization capabilities (Miettinen et al., 2014).

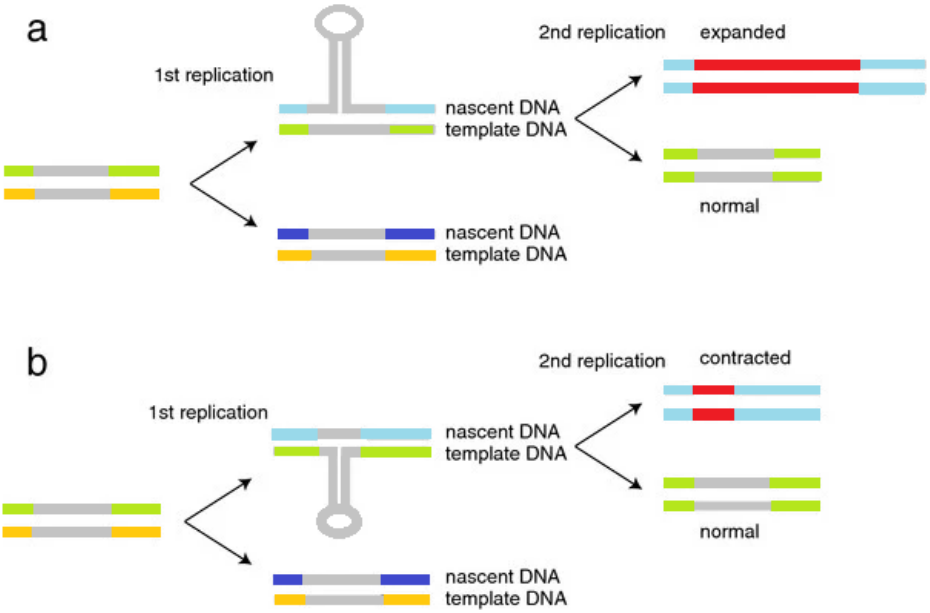

Figure 4: Hairpins in replication model. (a) Hairpin in nascent DNA leads to expanded TNR in one chromatid. (b) Hairpin in template DNA results in under-replication of a segment in one chromatid (image adapted from Liu et al., 2010).

1.1.4 Non-B DNA structures, R-loops and instability of CAG repeats

A B-DNA structure is the right-handed DNA helix form of DNA which is the canonical structure (Potaman & Sinden, 2013). Non-canonical structures or non-B DNA are another structural form DNA presents. Some non-B DNA structures include: cruciforms, slipped structures, triplexes, G-quadruplexes and Z-DNA (Zhao et al., 2010) (figure 5). These particular structures have been implicated in problems with DNA replication and transcription ultimately contributing to genome instability (Bacolla et al., 2004; Bacolla & Wells, 2004; Wang & Vasquez, 2006).

Exposure of single-stranded DNA regions to replication, repair or recombination allows for formation of non-B DNA structures in repeat containing regions (Lin et al., 2010). One formation that can arise from exposure is called R-loop. A R-loop is a three-stranded nucleic acid structure, composed of a DNA:RNA hybrid and an exposed/displaced strand of DNA (Hegazy et al., 2020) (figure 6). R-loop structures are of important mention since they tend to form in GC-rich DNA because of guanine and cytosine stability and the higher the GC-content in a gene the easier its R-loop formation (Roy & Lieber, 2009; Sugimoto et al., 1995). According to Lin *et al.*, it was demonstrated that R-loops can

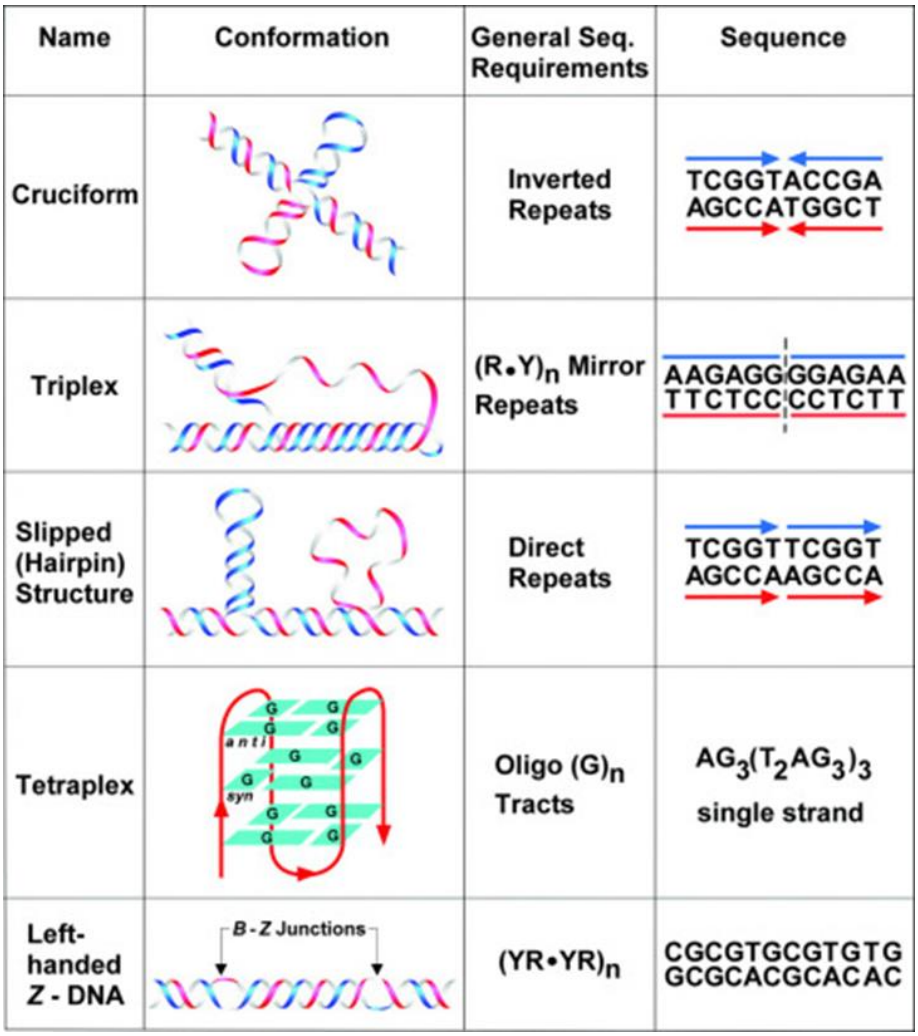

Figure 5: Non-B DNA structures with requirements and sequence examples (image from Bacolla & Wells, 2004).

indeed, form in CAG repeat tracts therefore having some degree of impact in polyQ diseases by generating DNA instability.

1.1.5 Chromatin changes influence instability in polyglutamine diseases

DNA methylation has been also shown to play a part in polyQ diseases. It has been demonstrated in mouse models that abnormal DNA methylation and histone H3 lysine 9 methylation occur at CpG islands around the expanded CAG tract when there is a defficiency in methyltransferases (Dnmt1 in this case) (Dion et al., 2008), showing a connection between chromatin management and repeat instability.

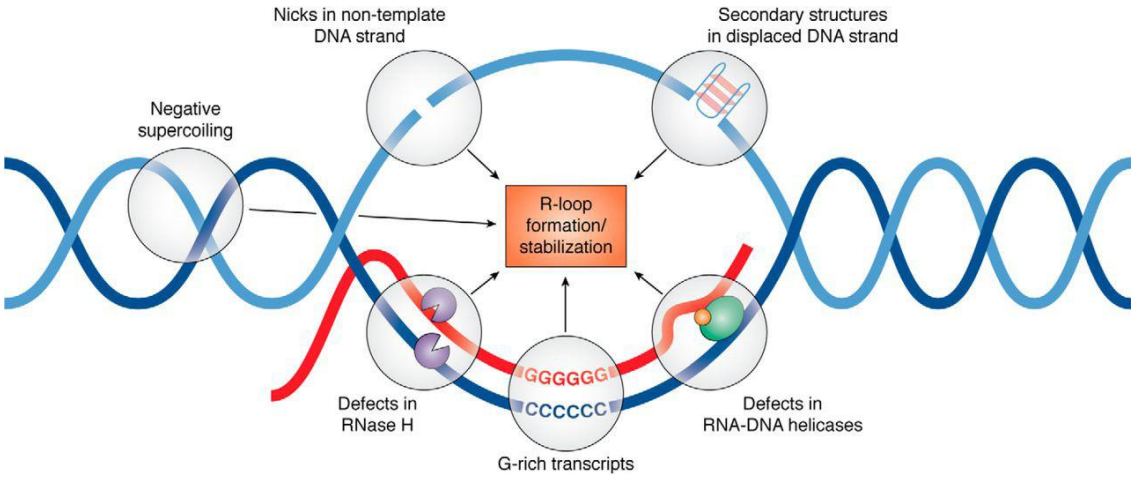

Figure 6: Schematic representation of factors promoting R-loop formation and stabilization (image from Hegazy et al., 2020).

Another epigenetic evidence was seen in transgenic mice carrying mutations of transcriptional repressor CTCF 11-zinc finger protein or CCCTC-binding factor (CTCF). CTCF could not bind its target sequence, and since its binding was compromised, it led to somatic repeat instability. This instability was also shown to be a product of DNA hypermethylation around the CTCF binding site at CpG loci (Libby et al., 2008).

1.1.6 Mutant proteins with glutamine-expanded tracts disrupt various downstream cellular pathways

With an increase of mutated proteins inside the cell, many functions can be compromised. Either the normal functions the wild-type protein would perform may not being executed as supposed or the mutated proteins may interact with new cellular pathways abnormally disrupting them. Additionaly, both scenarios may also take place. Table 2 shows a list of the cellular mechanisms (both nuclear and cytoplasmatic processes) that are most affected in HD, according to various literature sources and presented by Lieberman et al., 2019.

Gene expression and other nuclear processes are highly compromised in the case of polyQ diseases admitedly by the simple fact that most proteins that are mutated in these disorders are directly involved in regulation of gene expression by interacting with transcription factors or being transcriptor factors themlseves. Other factor is related to the nuclear inclusions (NIs). NIs directly compromise nuclear processes by sequestering important elements (including transcription factors (TFs)) unabling them to properly function.

Table 2. Cellular processes affected and implicated in Huntington's disease neurodegeneration, according to Lieberman et al., 2019.

Processes	**Mechanisms**
Nucleus and gene expression	Transcriptional dysregulation
	DNA damage and repair
	Impaired nucleocytoplasmic transport
	Irregular mode of mRNA translation (RAN translation*)
	RNA toxicity
Cytoplasm processes	Excitotoxicity
	Abnormal intracellular Ca2+ buffering
	Mitochondrial dysfunction
	Autophagy dysregulation
	Astrocyte dysfunction
	Altered axonal transport
	Neurotrophic factor dysregulation

Caption: RAN translation: Repeat Associated Non-AUG translation, or RAN translation, is an irregular mode of mRNA translation that can occur in eukaryotic cells.

Examples of imprescindible factors that, when tampered with, contribute to dysregulation of transcription are cAMP response element-binding (CREB) proteins, TAFII130 and specific-protein 1 (Sp1), which are directly linked to histone acetylation (Dunah et al., 2002; Nucifora et al., 2001). Besides transcription, splicing may also be affected as a consequence of CAG expanded RNAs. There is evidence that expanded RNAs sequester a splicing factor called muscleblind-link-1 (Mbnl1) and causes splicing alterations (Mykowska et al., 2011).

On a cytoplasmatic level, many problems can arise from the generation of abnormaly expanded proteins and the aggregates they may form over time. Axonal transport is one of the mechanisms that mat show problems. A good example is mutated huntingtin (Htt). Normal Htt is partly responsible by the motor machinery acting as a scaffold protein in multiprotein complexes (Saudou & Humbert, 2016). Reduced levels of normal function

Htt lead to abnormalitites to fast axonal support (FAT), in other words, makes FAT slower and inefficient (Lieberman et al., 2019).

Another mechanism present across most polyQ diseases is mitochondrial dysfunction. It is belived that mitchondria impairments/dysfunctions result from compromised and dysregulated transcription as was described above (Cui et al., 2006).

Since the type of cells most affected are neuronal cells, excitotoxicity and ion homeostasis dysregulation are highly taken into account and investigated as having a role in polyQ disease pathophysiology. It was demonstrated that expanded polyQ proteins, namely Htt, can increase ion influx to the cell by tampering with ligand-gated ion channels such as N-methyl-D-aspartate (NMDA) receptors. A study showed that increased levels of mutated Htt enhanced the activity of the NR2B subtype of NMDA receptors, leading to the increase of Ca^{2+} passing through these receptors (Raymond, 2017). With the increased influx of Ca^{2+}, neuronal dysfunction increased aswell.

1.1.7 Protein aggregation and inclusion formations in polyglutamine diseases

Mutated proteins derived from CAG TNR expansions in coding parts of their respective genes can aggregate and form oligomers of glutamine-expanded proteins. Stretches of polyQ are able to form hydrogen bonds to other long stretches of polyQ in an antiparallel fashion (Perutz et al., 1994). Aggregates formed by this hydrogen bond linkage between glutamine stretches may further join with other molecules that are recruited to form β-pleated sheets leading to increased aggregation and precipitation within the cells (Ross, 1997). Mutated protein aggregates/oligomers are found within cytoplasm of neuronal cells and sometimes in other type of cells. These aggregates can sometimes translocate to the nucleus, a process that highly relies on protein structure and size. Smaller proteins, like ataxin-3, may be able to enter the nucleus through passive diffusion (Paulson et al., 1997). One other very important factor for nuclear translocation is the existence of a nuclear localization signal (NLS) in the protein structure. Proteins that are too big or do not have an NLS need to be cleaved by proteases to a certain point to be able to enter the nucleus by diffusion (DiFiglia et al., 1997; Goldberg et al., 1996) or they can only enter by active translocation.

1.1.8 Neuropathological findings in polyglutamine diseases

The molecular mechanisms of neurodegeneration that underly polyQ diseases are not yet fully understood. It is believed that mutant proteins undergo conformational changes due to the abnormal number of glutamines, impacting their normal functions and turning them more prone to degradation (Ordway et al., 1997). The cell cannot fully destroy these altered-function proteins completely and the inefficient cleavage by caspases plus labelling with ubiquitin often leads to aggregation with other important cell elements such as TFs, chaperones and proteosomes (Tunalı, 2020). Depending on the subcellular distribution of the protein and the regions where it is mostly expressed, certain areas begin to show aggregates within the cell and/or nucleus.

In most polyQ diseases, selected brain areas are affected and show these aggregates in histological cuts of post-mortem patients' tissue. In SBMA, aggregates of expanded protein are also found within skeletal muscle showing that these aggregates are not exclusively found in the brain (Morishima et al., 2008). Brain areas usually affected in HD are the striatum, the dentatorubral system (dentate nucleus to red nucleus), in DRPLA and the cerebellum and pons, in the SCAs. Substantia nigra is also affected in both HD and SCA3. In the long term, aggregates in these brain regions interfere with the normal neuronal functioning. With insistent cellular interferences, neurons eventually degenerate leading to neuronal loss in affected brain regions. Figure 7 shows examples of protein aggregates and inclusions observed on both mouse and human samples.

1.1.9 Most evidential symptoms across polyglutamine disorders

The most prevalent and known polyQ disease is Huntington's disease. One of the most predominant symptoms of this disorder is chorea. Chorea in Huntington's is so characteristic that the disease is often referred to as Huntington's chorea. Another symptom well documented for some polyQ diseases is ataxia. Across all 6 SCAs of polyQ cause, ataxia is present with great similarities because disease advance leads to cerebellum lesion/malfunctioning and degeneration (Paulson, 2009).

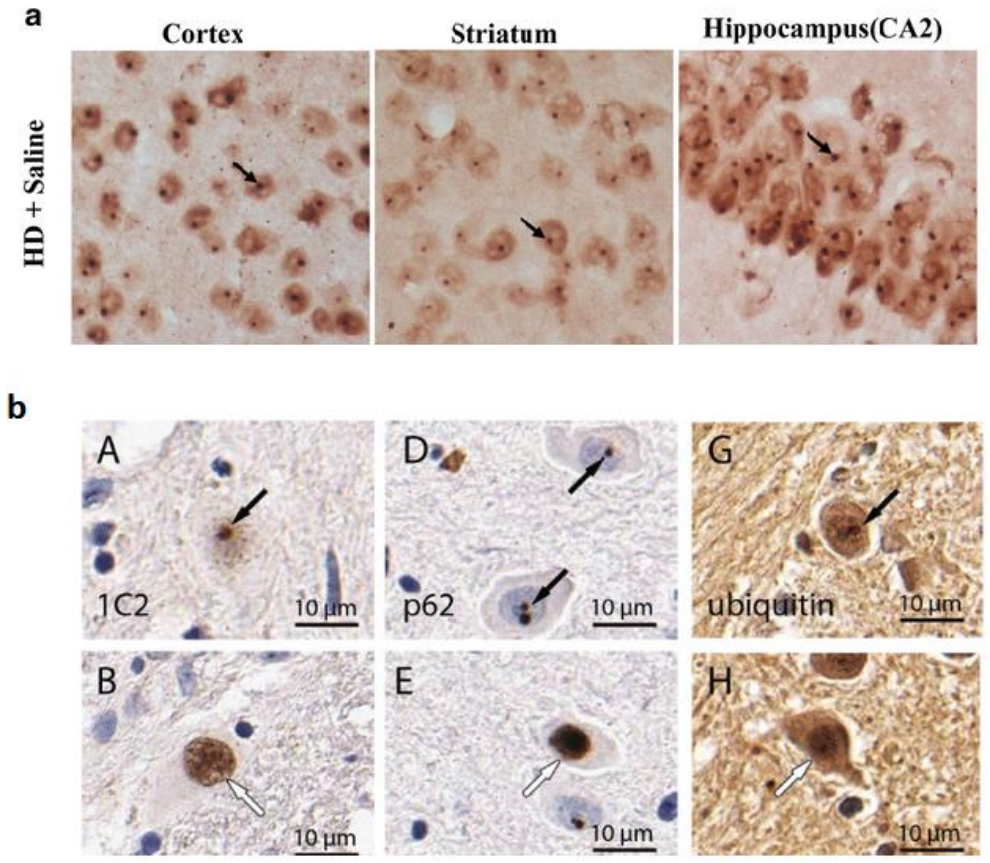

Figure 7: a - Protein aggregates in Huntington's disease mouse tissues (cortex, striatum and hippocampus) marked by black arrows. b – Aggregates and inclusions in SCA3 affected patients' lower pons tissue. A-B: marked with 1C2, an anti-PolyQ-expansion diseases marker antibody; D-E: marked with p62 an autophagy marker; G-H: marked with ubiquitin. Black arrows in A, D and G point to nuclear inclusions. Image segments from Singh et al., 2018 (a) and Seidel et al., 2012 (b).

Similar symptoms of all polyQ diseases include dementia, muscle weakness, epilepsy/seizures and cognitive impairments. It is known that some diseases in this group may show these symptoms more profoundly than others, but this similarity of symptoms can often hinder the diagnose if looking at the symptoms alone. Fortunately, with nowadays knowledge of these 9 diseases, including their major effects on the peripheral and central nervous system, along with the available genetic testing tools, it is possible to reach a clear diagnose/conclusion faster than ever and discuss therapeutic approaches (Huang et al., 2019).

1.1.10 General epidemiological features of polyglutamine diseases

PolyQ disorders affect and can occur in all population groups and cases are seen through the entire globe. However, in some cases, particular polyQ diseases can present themselves only in particular areas of the world, which leads to the belief that founder effects may have occurred. In the field of Population genetics and according to the National Human Genome Research Institute, founder effect is defined as a *genomic variability that occurs when a small group of individuals becomes separated from a larger population*, in other words, a genetic alteration observed with high frequency in a group that is or was geographically or culturally isolated, in which one or more of the ancestors was a carrier of the altered gene. This can result in increased frequency of inherited diseases in limited population groups and is observed in polyQ disorders with regional variability.

From all the glutamine-expansion diseases, Huntington's disease is the most common among them. Around 1 in 10.000 persons are affected and although it is observed around the globe, it is more prevalent in Canada and in the USA (Rawlins et al., 2016).

SBMA or Kennedy disease is a very rare disease with X-linked inheritance and its prevalence sits at around 1 in 300.000 males (Stoyas & La Spada, 2018). This condition has been studied mainly in Japanese males (Tanaka et al., 1996) where it is believed the prevalence of the condition is due to founder effects, but several European countries also have reported cases of SBMA.

DRPLA is another very rare polyQ disease that occurs predominantly in the Japanese population and its prevalence is 0.48 in 100,000 individuals (Tsuji et al., 2008). DRPLA has also been identified in other populations in European and North and South America (Veneziano & Frontali, 2016).

SCAs are a group of more than 40 diseases (Sullivan et al., 2018), but only 6 can be grouped in polyQ diseases as mentioned previously (other SCA types are not caused by CAG expansions, or are caused by expansions in non-coding gene regions). The overall prevalence of autosomal dominant ataxia (ADCA) is estimated at 1:37.000 individuals worldwide, according to Orphanet. SCA1 makes about 6% of all ADCA (Opal & Ashizawa, 2017), SCA2 takes about 13% (Pulst, 2019) and SCA3 is the most common ADCA in most populations, but with no concrete prevalence values worldwide

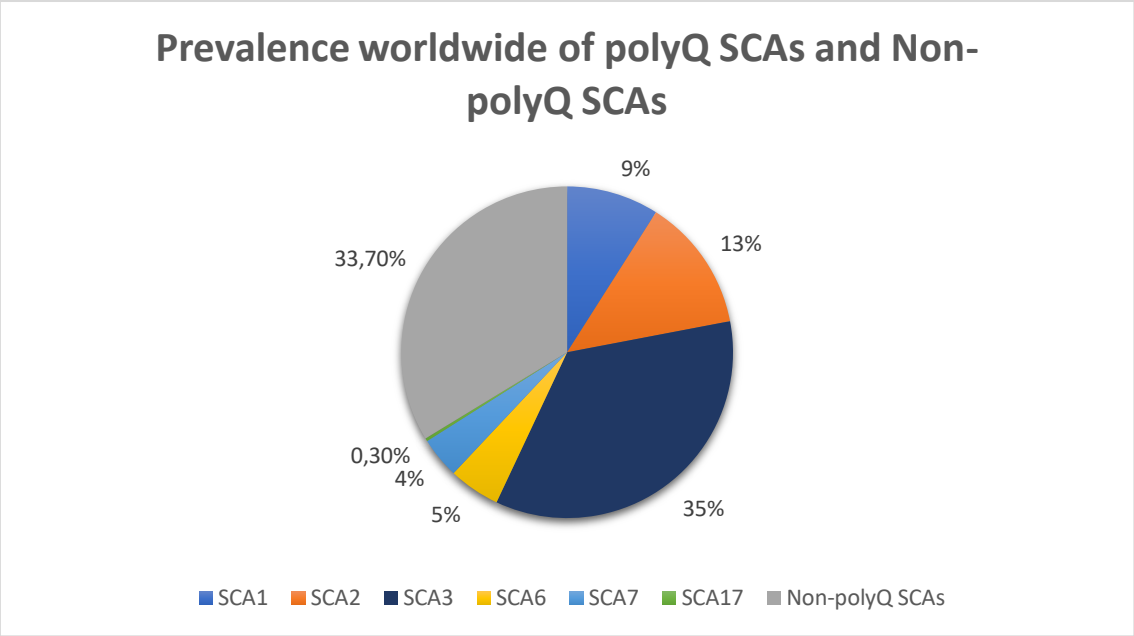

Figure 8: Pie chart with polyglutamine SCA prevalence in ADCAs.

(Paulson & Shakkottai, 2020; Stoyas & La Spada, 2018). The prevalence of SCA6 is estimated at less than 1 in 100.000 (0.02:100.000 to 0.31:100.000) and according to the last update of Casey & Gomez, 2019 fewer than 10.000 individuals were identified with SCA6. SCA7 is another less common ADCA with its prevalence sitting around 1 in 500.000 individuals and it accounts for around 2% in overall SCAs (Filla et al., 2000; Tang et al., 2000). Finally SCA17 is by far the most rare disease included in polyQ ADCAs, with less than 100 families reported and with a prevalence of 0.3% in ADCAs (Maruyama et al., 2002; Stoyas & La Spada, 2018; Toyoshima et al., 2022).

Figure 8 groups, in a pie chart, the estimated prevalence of every polyQ SCA and non-polyQ SCA prevalence in overall SCA prevalence worldwide, according to the previous epidemiological information.

1.2 Characteristics of individual polyglutamine diseases

PolyQ diseases have similar characteristics when it comes to disease mechanism, type of mutation and pathological hallmarks. Despite these similarities, the nine disorders are separate pathological entities, with many features that set them apart. This topic will further elucidate and give information on some polyQ diseases as unique conditions.

HD and SCA1 are not included here, since the information about them is later mentioned in the results as examples of the data presented for each disease in the website.

1.2.1 Dentatorubral-pallidoluysian atrophy

The first description of dentatorubral-pallidoluysian atrophy (DRPLA) was in 1946, with two cases in a family that showed symptoms assumedly related to a certain type of neurodegenerative disease (Titica & Van Bogaert, 1946). A few years later (1958), a study presented a case of this disease that remained unidentified for a few more years (Smith et al., 1958). DRPLA was given its name in 1982 by two scientists (Naito & Oyanagi, 1982).

As an autosomal dominant inherited disease, DRPLA is passed on to offspring of an affected individual with a 50% chance and as it happens with many other glutamine expansion disorders, the severity and precocity of the age of onset is proportional to the size/number of triplets inherited (higher the number of triplets, earlier disease onset and more severe cases).

The triplet expansion occurs in the *ATN1* gene located in chromosome 12. According to GeneReviews, affected individuals tend to have 48 or more CAG repeats within exon 5, whether in normal individuals the number of repeats ranges between 6 and 35 (Veneziano & Frontali, 2016).

The *ATN1* gene codes for the protein atrophin-1. Atrophin-1 is part of a group of conserved transcriptional corepressors involved in nuclear transcription and signalling mechanisms. From its domain structure (figure 9), it is suggested that atrophin-1 acts both in the nucleus and the cytoplasm along with other specific regions such as perinuclear regions and cell junctions. This protein has both a nuclear localization signal (NLS) and a nuclear export signal (NES) and has domains for interacting with specific proteins that are directly involved in nervous system development (Yazawa et al., 1995).

Pathophysiological mechanisms of DRPLA have been described and revolve around accumulation of mutated protein and inability to degrade it. According to S.Tsuji this accumulation was shown to be more diffuse inside cells and not seen as much inside the nucleus as NIs (Tsuji, 2011). In 2014, a study presented evidences that epigenetic

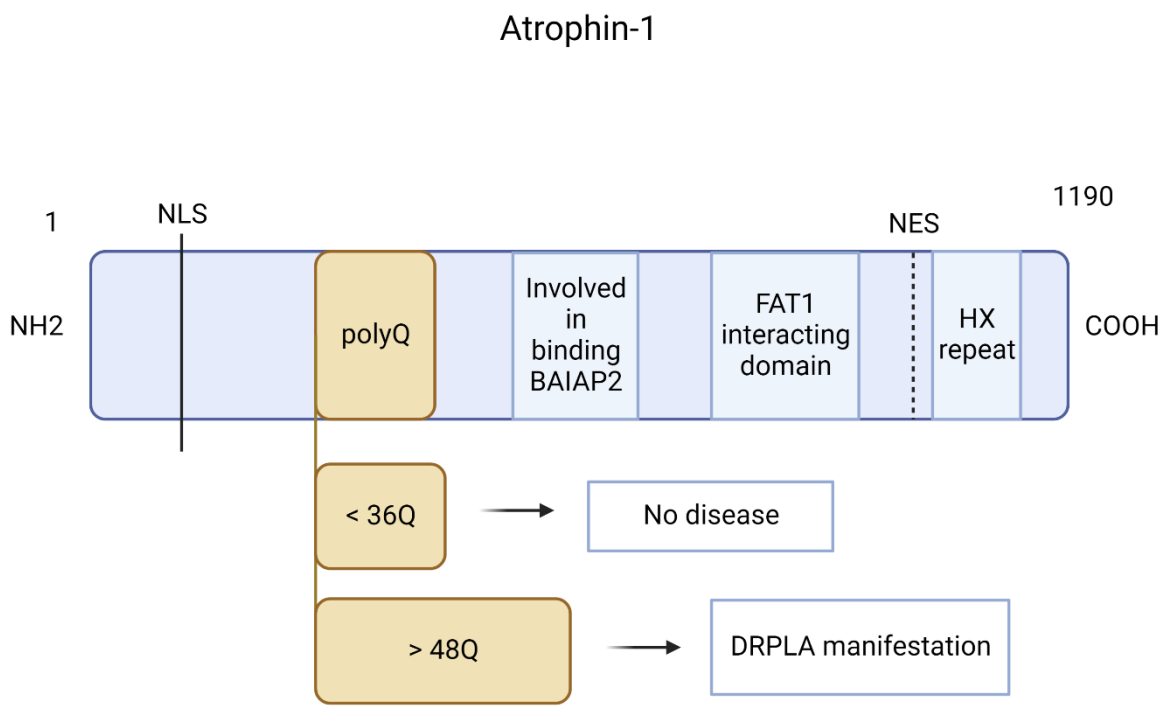

Figure 9: Atrophin-1 protein structure and important domains along with range of glutamine in expanded and normal cases. BAIAP2: BAR/IMD domain containing adaptor protein 2; FAT1: FAT atypical cadherin 1; HX repeat: Hemopexin-like repeats

modulation of the *ATN1* constitutes a very important regulation mechanism for proper functioning and development of the nervous system (F. Zhang et al., 2014). It was mentioned that a lysine-specific demethylase 1 (LSD1) targets *ATN1* and further enhances cell development in the cortex. With increased concentration of *ATN1* inside the cell, the more demethylation of it will occur and this, in turn, will continuously increase neuronal development in the cortex. Added to the neuronal expression dysregulation, the toxic levels of atrophin-1 translated inside will also increase. In DRPLA, cells are unable to destroy mutated atrophin-1 proteins either because of conformational changes or compromised cell clearance tasks (F. Zhang et al., 2014).

Although atrophin-1 interacting epigenetic factors show crucial changes that tampers cell function, another very important pathophysiological mechanism is related to transcription. As mentioned before atrophin-1 is a transcriptional corepressor and interacts with transcription factors within the nucleus. Mutated atrophin-1 undergoes conformational changes because of the abnormal number of glutamines in its structure, this affects binding to its regular "partners" and creates abnormal binding to other elements that are not usually bound to atrophin-1 in normal circumstances. A study in 2005 presented that mutated atrophin-1 could bind TAFII130 and totally deprive it of its

transcriptional functions and besides disrupting TAFII130 functions, it also disrupted CREB-dependent activated processes, namely *c-fos* activation and suppression (Shimohata et al., 2005).

1.2.2 Spinal and bulbar muscular atrophy

Spinal and bulbar muscular atrophy (SBMA) is a peculiar condition within polyQ diseases as it affects both the nervous system and the hormonal balance, in the more severe cases. SBMA had its first evidential descriptions in Japanese literature in 1897 by Hiroshi Kawahara (Takahashi, 2001). This condition was only fully described as a new type of disease and named SBMA in 1968 by William Kennedy which is why literature may also present this disease as Kennedy's disease (Kennedy et al., 1968).

SBMA is a disease with recessive inheritance patterns linked to the X chromosome, as mentioned earlier. Females who have the mutation on one of the X chromosomes are considered carriers but in some more extreme cases they can also present mild cases of SBMA even with a "backup" X chromosome. Males are more prone to show the full effects of the mutated gene since they do not have a second X chromosome without the mutation as failsafe, therefore male offspring are mainly affected. In simple terms, the Muscular Dystrophy Association presents that: each son of a woman who carries an X-linked disease has a 50 percent chance of inheriting the gene flaw and developing the disease. Each daughter has a 50 percent chance of inheriting the gene flaw and being a carrier herself. Throughout the years, SBMA failed to be properly diagnosed as the symptoms presented by affected individuals are often similar to those of other neuromuscular disorders such as amyotrophic lateral sclerosis (ALS). Today SBMA is less misdiagnosed due to genetic testing.

In SBMA, the *AR* gene has expanded CAGs in exon 1. The normal number of CAG triplets ranges until 34, penetrance increases between 35 and 46 repeats and above 47 repeats there is nearly 100% risk of developing disease in affected males (Laskaratos et al., 2021).

The *AR* gene codes for the androgen receptor (AR) which is a type of nuclear receptor that is activated by binding any of the androgenic hormones (testosterone is an example). It is involved in various cellular and biological functions, being the most important ones cell and organ development, intracellular response to hormones and the role as a

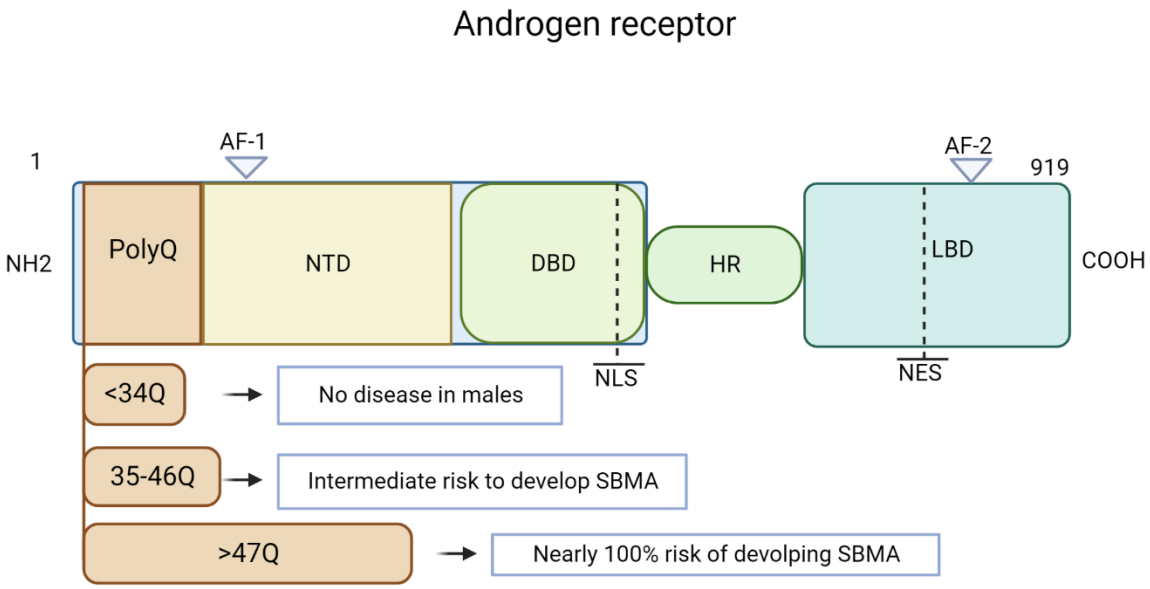

Figure 10: Androgen receptor scheme with important domains and regions represented. NTD: N – terminal transactivation domain; DBD: DNA binding domain; LBD: Ligand binding domain; HR: hinge region.

transcription factor in AR signalling pathways (Hsiao & Chang, 1999; Wissmann et al., 2007). AR has very important conserved interacting domains along with an NLS and a NES (figure 10). AR is found in the cytoplasm when unbound and in the nucleus when bound to one of its many ligands ultimately altering gene expression of its target genes.

SBMA disease progression is mostly related to loss of receptor sensitivity and function. Glutamine-expanded receptors tend to be less active and sensitive and this loss of sensitivity is proportional to the length of the glutamine tract (Lieberman, 2018). This loss of sensitivity impairs the AR signalling cascade, which consequently dysregulates the transcription of AR responsive genes.

Mutated receptors can sometimes bind its ligands (even with expanded glutamine tracts) and be translocated to the nucleus where it unfolds and aggregates.

Another pathophysiological mechanism that has been suggested revolves around increased mitochondrial impairment and autophagy. There has been evidence that polyQ-AR led to the accumulation of microtubule-associated protein 1 light chain 3 (LC3), an autophagy marker associated with increased autophagosome numbers, and aggregation (Cortes & La Spada, 2015).

1.2.3 Spinocerebellar ataxia type 2

Spinocerebellar ataxia type 2 (SCA2) is a condition characterized by progressive problems with movement and is one of the most common and severe SCAs. SCA2 is part of the ADCA I group of diseases. ADCA I diseases involve cerebellar ataxia along with variable pyramidal, extrapyramidal, and neuropathic signs. As other ADCAs, SCA2 is inherited in an autosomal dominant manner. This means that having one mutated copy of the responsible gene in each cell is enough to cause signs and symptoms of the condition.

First SCA2 descriptions were made by Wadia and Swami in a 1971 study where authors presented the early appearance of slow saccadic eye movements and peripheral neuropathy in patients. The condition was believed to be a unique form of autosomal dominant ataxia (Wadia & Swami, 1971).

The mutation in SCA2 is found in the *ATXN2* gene (chromosome 12) most precisely in its first exon. Normal range of CAG repeats in this exon do not go past 31/32, but in SCA2 scenarios, the range goes beyond 32. SCA2-causing alleles have 33 or more CAG repeats and the most common disease causing number of repeats ranges between 37 to 39 (Pulst, 2019).

ATXN2 codes for the ataxin-2 protein. Ataxin-2 is a 1313 amino acid protein with 140.283 Da protein, found mostly expressed in the brain, heart, liver, skeletal muscle, pancreas and placenta. Isoform 1 is predominant in the brain and spinal cord, while isoform 4 is more abundant in the cerebellum. In the brain, ataxin-2 is broadly expressed in the amygdala, caudate nucleus, corpus callosum, hippocampus, hypothalamus, substantia nigra, subthalamic nucleus and thalamus. This protein is involved in RNA binding and other cellular processes. It has been reported that ataxin-2 is important for normal epidermal growth factor receptor (EGFR) trafficking by negatively regulating the receptor internalization (Nonis et al., 2008) and it participates in transport and metabolic processing of RNA (Castello et al., 2012). Another important cellular target of ataxin-2 is protein-c (Ralser et al., 2005). Lastly there has been evidence of ataxin-2 having an important role in translation regulation (Satterfield & Pallanck, 2006) and in P-body/stress granules assembly (Nonhoff et al., 2007). Ataxin-2 is mainly localized to the cytoplasm and semi-quantitative immunofluorescent analysis of normal brains has revealed that cytoplasmic ataxin-2 strongly associates with ribosomal protein S6 suggesting an important role of ataxin-2 in protein synthesis (Watanabe et al., 2020) and

agreeing with what was mentioned earlier. Ataxin-2 structure is presented in figure 11 with all the important interacting domains and their approximate positions.

The CAG-triplet expansion in the ATXN2 gene gives rise to an abnormal polyQ region in the ataxin-2 protein, which results in a gain of toxic function that affects specific groups of neurons (Lastres-Becker et al., 2008; Velázquez-Pérez et al., 2011). Glutamine expanded ataxin-2 undergoes a conformational shift to a β-sheet–rich structure, which forms insoluble aggregates with β-sheet–rich amyloid fibrillar structures and accumulates as inclusion bodies in neurons. The aggregation of these forms triggers a series of events that lead to programmed cell death and consequently degeneration of neuronal structures (Matilla-Dueñas et al., 2009). Studies have presented evidence that polyQ expansion in ataxin-2 lead to abnormal autophagy and enhanced expression of a multifunctional protein involved in regulating RNA metabolism called Staufen1 (STAU1) (Paul et al., 2018). Mutated ataxin-2 increases STAU1 stability when bound to it, therefore increasing its levels. Aberrant levels of STAU1 induce uncontrolled RNA processing and compromised mRNA transport (Paul et al., 2018). As seen in other diseases caused by polyQ expansion, mitochondrial stress is also observed in SCA2, with increased superoxide dismutase (SOD) levels and decreased catalase expression. A study revealed the increased levels of the intracellular hydrogen peroxide, produced by the increased SOD from superoxide, could not be decomposed by the reduced levels of catalase, thus

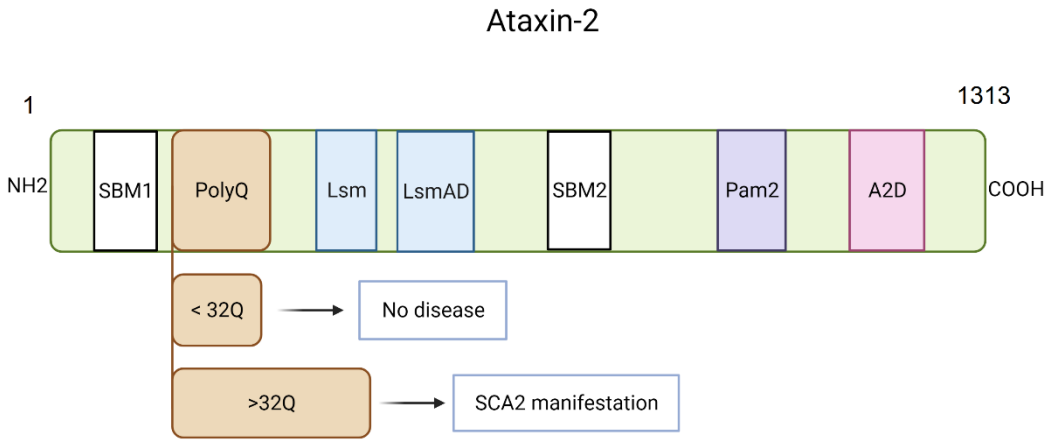

Figure 11: Ataxin-2 structure. SBM1/2: SRC homology 3 domain binding motifs 1 and 2; Lsm: Like-Sm protein; LsmAD: LsmAD domain-containing protein; Pam2: PABP-interacting motifs; A2D: Ataxin-2 domain protein; PolyQ: polyglutamine tract.

leading to increased oxidative stress, disturbances in the antioxidant system, changes in the oxidative phosphorylation system, and abnormalities in the mitochondrial activity that were detected in the fibroblasts from SCA2 patients (Cornelius et al., 2017).

1.2.4 SCA3

Spinocerebellar ataxia type 3 (SCA3), also known as Machado-Joseph disease (MJD), is an inherited disease of the central nervous system involving progressive ataxia. It results from a specific genetic defect that leads to impairment of nerve cells in the brain and nerve fibers carrying messages to and from the brain (Paulson & Shakkottai, 2020). SCA3 is part of the ADCA I group of diseases along with SCA1, SCA2, SCA17 and many others, all of which are inherited in an autosomal dominant manner meaning that each child of a heterozygous affected individual has a 50% chance of inheriting the pathogenic mutation. The symptoms of SCA3/MJD usually appear in middle adult life and progress over several decades, with some patients surviving for almost 30 years after the onset of symptoms. Onset in adolescence or as late as age 70 can also occur. This extreme range in age of onset reflects differences in the size of the disease-causing DNA repeat (Paulson, 2009).

MJD was originally described in members of the families of Machado, Thomas, and Joseph originating from São Miguel Island, Azores, Portugal, in 1972 (Meira et al., 2020). In 1980, it was named Machado-Joseph disease, honouring the first patients and descendants who suffered from the disease (Lima & Coutinho, 1980), but only in 1983 it was first described as SCA3/MJD by researchers.

The gene associated with SCA3/MJD is denominated *ATXN3*. It has a size of around 42 kb and it is located on chromosome 14, cytogenic band 14q32.12, with a minus strand orientation. The gene has 11 exons, according to many sources and the pathological CAG repeats are found within exon 10 (Bettencourt & Lima, 2011). The repeat length of the SCA3 gene in healthy individuals varies between 12 and 44 trinucleotides while SCA3 patients have at least one allele with a CAG repeat tract within a range of 53 to 87 units. Expanded SCA3 alleles display intergenerational instability with a tendency for further expansion.

The *ATXN3* gene codifies a protein named ataxin-3. Ataxin-3 has 361 amino acids (in its isoform 2), 41250 Da, and is widely expressed throughout the organism (Evers et al., 2014). An approximate representation of the important known domains/motifs of ataxin-3 is shown in figure 12. This protein is admitted to be involved in the cellular clearance processes. It is suggested that it plays a role in the destruction and excess removal of damaged proteins through ubiquitin-proteasome system (UPS) by modulating polyubiquitination signals through its enzymatic activity as a deubiquitinase (Laço et al., 2012) (Bai et al., 2013; Bernardi et al., 2013). Besides having a role in the UPS, it also is capable of binding ATPases (Q. Wang et al., 2006) and identical proteins acting as a transcriptional corepressor (Klockgether, 2007). There has been evidence that ataxin-3 is involved in nervous system development, actin, intermediate filament and microtubule cytoskeleton organization (Rodrigues et al., 2010) along with a role in chemical synaptic transmission (Schöls et al., 1995). Lastly, it has been demonstrated that ataxin-3 aids in nuclear excision repair by interacting with the two human homologs of yeast DNA repair protein RAD23, HHR23A and HHR23B (G. H. Wang et al., 2000).

SCA3 is admitted to result from misfolding and aggregation of ataxin-3 caused by expansion of the polyQ tract of the protein. There are various hypotheses when it comes to aggregation and formation of inclusions. These aggregates and formation of inclusions can be seen as a protective cellular mechanism (Arrasate et al., 2004; Ross & Poirier, 2004).

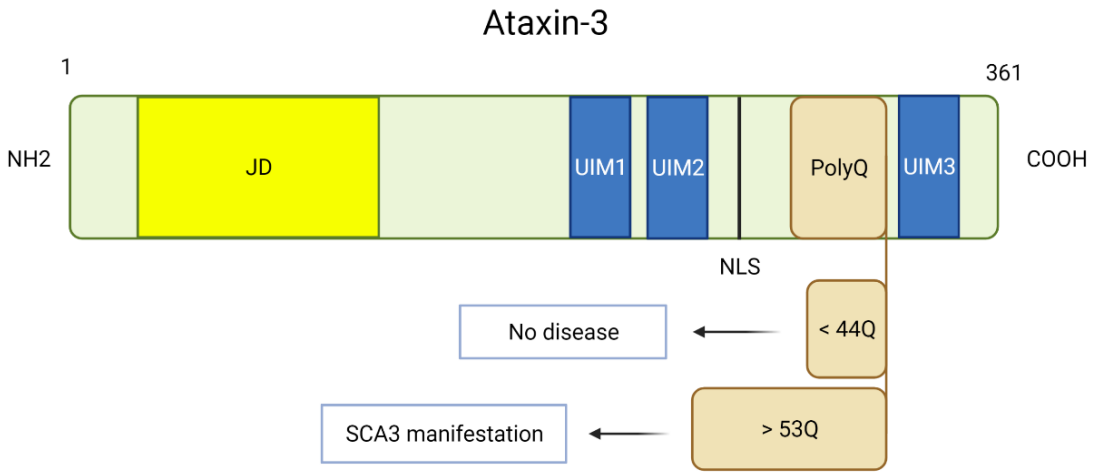

Figure 12: Ataxin-3 protein structure. JD – Josephin-domain of ataxin-3; UIM1-3: ubiquitin-interacting motif 1, 2 and 3; PolyQ: polyglutamine tract; NLS: nuclear localization signal/sequence.

On the other hand, these aggregates have been presented and implicated in further disease development since they are able to sequester other normal functioning proteins and compromise cell functions (McCampbell et al., 2000; McLoughlin et al., 2020; Olzscha et al., 2011; Paulson et al., 1997). Mutated ataxin-3 affects the protein quality control systems such as UPS, molecular chaperones and autophagy. These systems are overwhelmed with the increased levels of aggregate-prone proteins leading to altered proteins homeostasis (McLoughlin et al., 2020). Mitochondria function and calcium homeostasis are also compromised. Studies have shown that the expression of mutant ataxin-3 may lead to increased oxidative stress and promote cell death mediated by mitochondria (Chou et al., 2006; Tsai et al., 2004; Yu et al., 2009). It has also been predicted that mutant ataxin-3 alters calcium homeostasis by binding and interacting with targets with which normal ataxin-3 is not involved (gain-of-function mechanism) (X. Chen et al., 2008; Pellistri et al., 2013). These interactions often lead to intracellular calcium overloads which in turn can lead to cell death by cytotoxic mechanisms (Zhivotovsky & Orrenius, 2011).

1.2.5 Spinocerebellar ataxia type 6

Spinocerebellar ataxia type 6 (SCA6) is a disease that is included in the ADCA III group of disorders. ADCA III incorporates diseases that manifest as pure cerebellar ataxia and besides SCA6, it also includes SCA5, SCA11, SCA26, SCA30, and SCA31. It has been noted, however, that a small subset of patients of ADCA type III can present with non-cerebellar signs including mild neuropathy, pyramidal signs, or parkinsonism (Fujioka et al., 2013). As the other diseases of ADCA III group, SCA6 is an autosomal dominant disease, which means that individuals of either sex are equally likely to inherit the gene and develop the disease, that passes directly from one generation to the next without skipping generations. Each child of a heterozygous person with SCA6 has a 50 percent chance of inheriting the gene that causes SCA6.

SCA6 was first described as a polyQ disease in 1997, when the disease was first correlated with increased CAG triplets in the affected gene (Zhuchenko et al., 1997).

The mutation in SCA6 is found in the *CACNA1A* gene (chromosome 19), in the 47th exon. The normal range of CAG repeat numbers is usually between 4 and 18, but beyond 19

and until 33 (max recorded repeats of SCA6 patients) SCA6 manifests (Casey & Gomez, 2019).

The protein codified by the *CACNA1A* gene is the voltage-dependent P/Q-type calcium channel subunit α1A or CACNA1A. This channel is mostly expressed and present in the brain (nervous tissue). Voltage-gated Ca2+ channels (VGCC) are of extreme importance because influx of Ca^{2+} through these channels regulates various cellular processes, including neurotransmission and gene expression and therefore they have been identified as important elements in membrane excitability synaptic transmission and integration, as well as gene regulation (Miller, 1997), triggering synaptic transmission and neuronal plasticity (Goldin et al., 2000). The protein structure of this calcium channel comprises 4 domains (figure 13), but it is also known that the C-terminus of the α1A subunit can be independently translated to a new protein called α1A C-terminal (α1ACT). This protein can also interfere with nuclear processes and has been found to be implicated in the pathophysiological mechanisms of SCA6. CACNA1A is found in the cell membrane and the C-terminus of this channel may be translocated and found within the nucleus (Nóbrega & Pereira de Almeida, 2018).

Initial evidence concerning the pathophysiological mechanisms of SCA6 showed that polyQ expansion in the C-terminus of the α1A subunit caused ataxia by altering calcium channel function, although this information was however later described as not entirely correct by some authors (Saegusa et al., 2007; Watase et al., 2008). The α1ACT protein has been directly associated with toxicity within neuronal cells when mutated, in this case when bearing an abnormal number of glutamines. One of the first clues into the importance of this C-terminal protein came from a study from 2003. Authors demonstrated that C-termini containing Q28 were significantly more toxic than C-termini harbouring Q13in HEK cell cultures and that the toxicity was independent of the function of the full-length protein (Kubodera et al., 2003). Futhermore, like other polyQ diseases, the toxicity was highly dependent on the nuclear localization of the expanded protein. The α1ACT with polyQ expansion was found within cerebellar tissue and mainly localized to Purkinje cell nuclei, and this nuclear

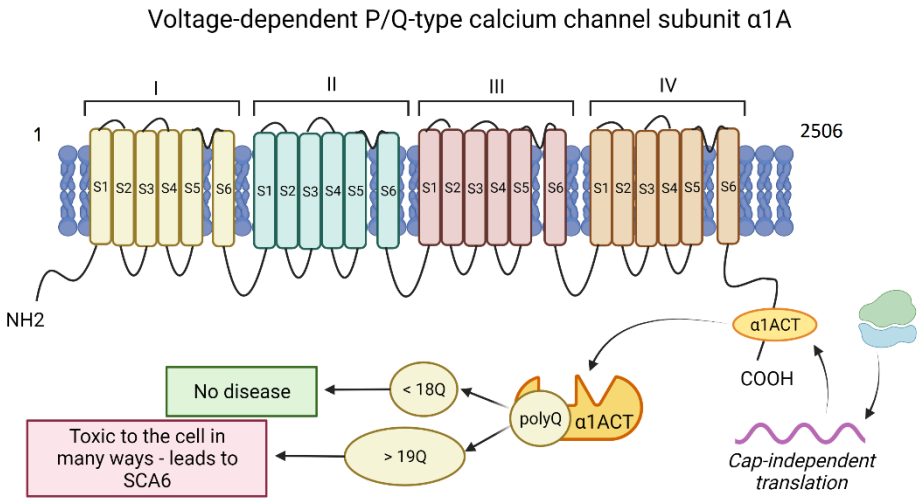

Figure 13: CACNA1A structure with the four transmembrane domains (I – IV) and their subdomains represented along with an approximated depiction of a1ACT containing the polyQ tract.

localization was not affected by polyQ length (Kordasiewicz et al., 2006). Compared to the properties of the normal α1ACT protein, the α1ACT with expanded glutamines has altered binding to various regular targets affecting normal cell functioning and leading to neuronal degeneration (Nóbrega et al., 2018).

1.2.4 Spinocerebellar ataxia type 7

Spinocerebellar ataxia type 7 (SCA7) is part of another group of diseases, set apart because of their exclusive disease mechanisms and the tissues affected: ADCA II. This group consists of diseases presented with cerebellar ataxia and retinal degeneration.

SCA7 was first described in 1937 and designated ADCA type II or olivopontocerebellar atrophy (OPCA) type III. It was distinguished from all others by the invariable presence of pigmentary retinal dystrophy. This subtype later became known as SCA7 (Froment J, Bonnet P, 1937).

The affected gene in SCA7 is *ATXN7* (located in chromosome 3). Wild-type alleles of this gene have between 4 and 36 CAG repeats, while mutated alleles have typically beyond 36 CAGs and can even reach more than 460 repeats. Among polyQ disorders,

SCA7 CAG repeats show the highest tendency to expand upon transmission, explaining the strong anticipation phenomena observed in families (Warrenburg et al., 2001).

ATXN7 codes for the protein ataxin-7. This protein is involved in transcription as an important component of co-activator complexes and is mostly expressed in the heart, brain, liver, and kidney. Ataxin-7 is involved and aids in transcription and microtubule skeleton stabilization processes: it may act as a subunit of General control non-depressible 5 (GCN5) and as a component of epigenetic control factors such as Spt-Ada-Gcn5 acetyltransferase (SAGA)-like complexes (Helmlinger et al., 2006), and it is necessary for microtubule cytoskeleton stabilization (Nakamura et al., 2012). Problems in ataxin-7 protein were shown to be directly involved in progressive visual impairments (Campos-Romo et al., 2017). Figure 14 presents a simplified representation of ataxin-7 and its important domains.

One major consequence of the polyQ expansion mutation of ataxin-7 is the intensive intracellular accumulation of the protein in tissues affected in the disease (ultimately leading to Nis). This accumulation of mutated ataxin-7 inside the nucleus is a consequence of caspase-7 mediated cleavage of the altered protein that generates a short fragment containing the amino-terminus with the polyQ tract but without the nuclear export sequence (NES). This fragment is short enough for passive diffusion through nuclear pore complexes but is retained in the nucleus since it lacks a NES. Within the nucleus, it may alter the function of SAGA complex, either by replacing the full-length protein in the complex or by inducing aggregation and sequestration of SAGA components. Accumulation of the mutant ataxin-7 fragment is associated with an increased acetylation at lysine-257 (K257) located close to the caspase-7 cleavage site (Mookerjee et al., 2009). In the absence of lysine acetylation, the fragment is degraded by autophagy. These results suggest that aberrant K257 acetylation prevents the clearance of the fragment by autophagy, and thereby slows down its turnover. The possibility that transcriptional alterations in SCA7 could result from dysfunction of SAGA acetylation and deubiquitination has been explored in several studies, especially because mutant ataxin-7 had been shown to properly incorporate into SAGA (Helmlinger et al., 2004; Palhan et al., 2005). The mechanism behind retinal degeneration revolves around GCN5. SAGA containing mutant ataxin-7 lacks critical subunits and leads to GCN5 or lysine

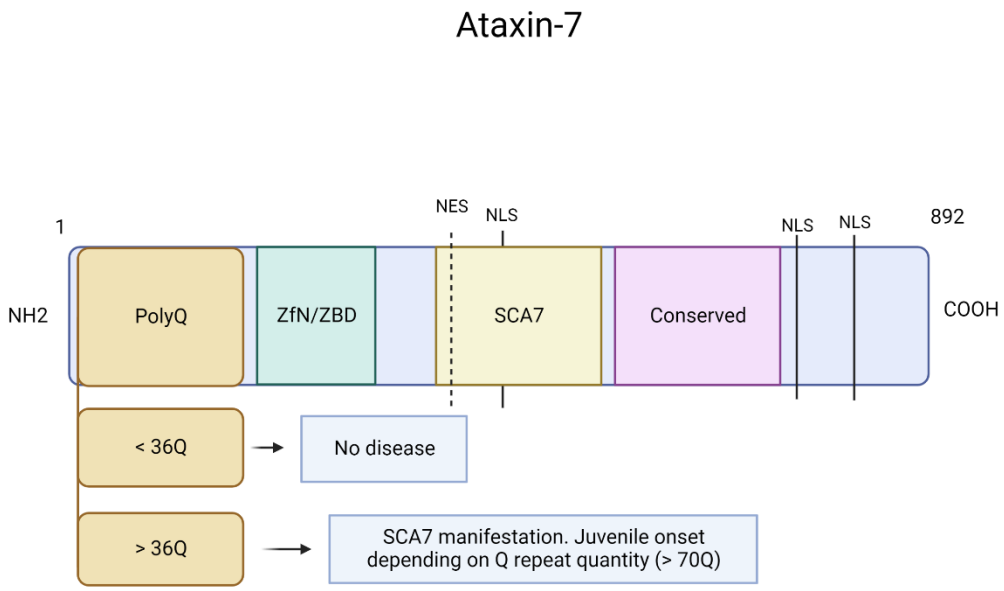

Figure 14: Ataxin-7 structure with various important protein interacting domains. Zfn: zinc-binding domain; SCA7: ATXN7 or SCA7 domain; Conserved: conserved domain.

acetyltransferase 2A (KAT2A) function problems. With GCN5 dysfunction, photoreceptor-specific genes have hypoacetylation of histone H3 which leads to decreased expression of these same genes (Palhan et al., 2005).

1.2.5 Spinocerebellar ataxia type 17

From all the polygQ diseases, Spinocerebellar ataxia type 17 (SCA17) is the one that was most recently described and is one of the most the rare conditions in glutamine expansion disorders and ADCAs. SCA17 is also part of the ADCA I. like SCA2 and SCA3 mentioned earlier.

In 1999, a polyQ expansion was identified in the transcription factor TATA-box binding protein (TBP) in a patient with ataxia with negative family history (Toyoshima & Takahashi, 2018). CAG/CAA repeat expansions in the *TBP* gene were identified in families with SCA, establishing this repeat expansion as the underlying mutation in the disease later classified as SCA17.

As mentioned before, in SCA17 the affected gene is the *TBP* gene and the pathological CAGs are found within exon 3 (McIntosh et al., 2021). Normal range to CAG repeats in

this exon tend to be between 25 and 42 but in mutated alleles this exon presents from 45 up to 66 repeats. This gene is located in chromosome 6.

TBP is a general transcription factor that binds specifically to a DNA sequence called the TATA box. This DNA sequence is found about 30 base pairs upstream of the transcription start site in some eukaryotic gene promoters. It functions at the core of the DNA-binding multiprotein factor, transcription factor II D (TFIID). Binding of TFIID to the TATA box is the initial transcriptional step of the pre-initiation complex (PIC), playing important roles in the activation of eukaryotic genes transcribed by RNA polymerase II (Hoffmann et al., 1990; Peterson et al., 1990). TBP also plays important roles in transcription mediated by RNA polymerase I and III, as it is a component of transcription factors that regulate transcription mediated by these forms of RNA polymerase (Friedrich et al., 2005; Gouge et al., 2015). TBP is solely found within cell nuclei. Figure 15 presents a simplified representation of TBP and its important domains.

Description of SCA17 pathophysiology and disease progression focus on two important pathological mechanisms: formation of nuclear aggregates or inclusions and reduction of endogenous normal TBP levels (Friedman et al., 2007; Huang et al., 2011).

Overexpression of full-length-mutant TBP and truncated-mutant TBP lacking the DNA-binding domains (DBDs) was found to cause formation of inclusions, suggesting that insoluble aggregates are disease factors and that the neurotoxicity of mutant TBP is independent of DNA binding (Hsu et al., 2014).

In mouse models, polyQ expanded TBP has been shown to bind less to DNA, and polyQ-expanded TBP fragments, which are incapable of binding DNA, formed nuclear inclusions and caused severe neurological phenotypes (Friedman et al., 2008). The same author also showed that truncated mutant TBP forms aggregate more easily than full-length mutant TBP in transfected cells which suggested that mutant TBP can induce neurotoxicity without binding to DNA and that truncated TBP is more toxic than full-length TBP (Friedman et al., 2007).

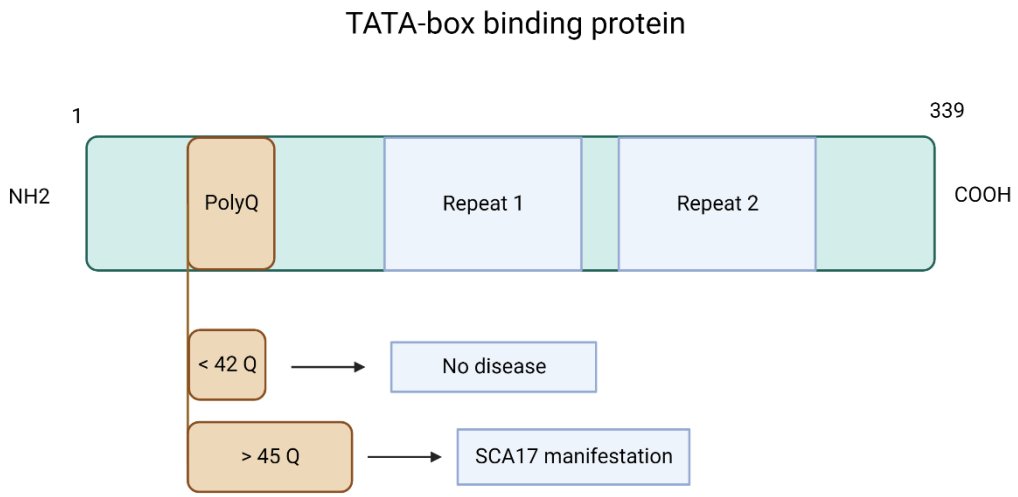

Figure 15: TBP structure with a representation of two repeats as a result of an imperfect direct repeat along with polyglutamine (PolyQ) tract location and following ranges in normal and disease scenarios.

1.3 What is a database?

The first database management system (DBMS) was first designed in 1960 by Charles Bachman. A simple definition of a database, offered by Oracle, is an organized collection of structured information, or data, typically stored electronically in a computer system with support for data manipulation. Examples of data that can be included in databases include raw text with various type of information (height, name, age, etc.), files, pdfs, pictures, images and more. Data is grouped in various types of data categories. There is textual data, numeric data, date and time, binary data and other data types (*Database Fundamentals #8: All About Data Types - DZone Database*, n.d.). Textual data is no more than a way to store words, numbers, symbols of different sizes and numeric data can only store numbers in integers, decimals and floats. Data and time category type is used to keep track of data and time information in order to be used in special calculations. When in need of storing documents (word documents for example) and programs, these are store as binary data. Finally, other data types category stores special types of information and data such as Extensible Markup Language (XML) and spatial information which stores either geographic or geospatial information so that you can define shapes, points, lines, etc. All this data and data types in the most common operating databases are sculpted in rows and columns inside tables to make querying and processing more efficient

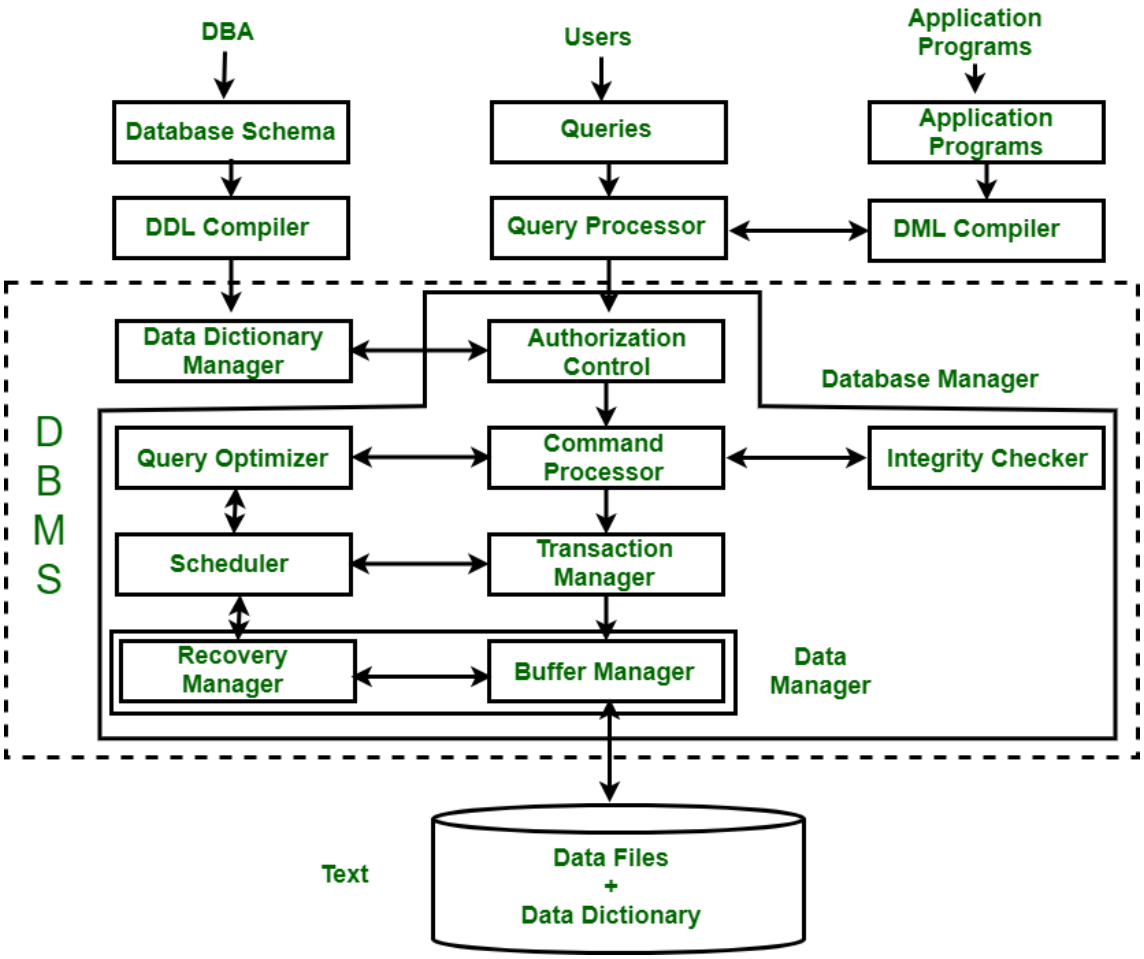

Figure 16: Database structure also known as overall system structure or database architecture and different from the tier architecture. Database structure/architecture is divided into three components: Query Processor, Storage Manager, and Disk Storage. DBA - database administrator; DDL – dynamic link library; DML - data manipulation language; (image from Geeksforgeeks.org, accessed online in: **https://www.geeksforgeeks.org/structure-of-database-management-system/**).

(*What Is a Database | Oracle Portugal*, n.d.). By being modelled this way, data can be easily accessed, managed, modified, updated, controlled, and organized by the DBMS.

Together, the data and the DBMS, along with the applications that are associated with them, are referred to as a database system, often shortened to just database. Figure 16 shows the overall database structure with all the components involved.

1.3.1 Database management system. What is it and what does it do?

A database management system (DBMS) is a system software (collection of programs) for creating and managing databases. In simple terms, a DBMS makes it possible for end users to create, protect, read, update and delete data in a database (*Database Management System (DBMS)*, n.d.). A DBMS essentially serves as an interface between databases and users or application programs, ensuring that data is organized and remains easily accessible.

A DBMS can be basically described as a computerized data-keeping system but, a more in-depth explanation of what a DBMS does reveals at least 9 important DBMS functions (*Functions of Database Management System (DBMS) - QS Study*, n.d.):

1. Data dictionary management
2. Data storage management
3. Data transformation and presentation
4. Security management
5. Multi-user access control
6. Backup and recovery management
7. Data integrity management
8. Database access languages and application interface
9. Database communication interface

1.3.2 Components of DBMS

A DBMS consists of multiple integrated components that deliver a consistent, managed environment for creating, accessing and modifying data in databases. Those components, according to Craig S. Mullins (*Database Management System (DBMS)*, n.d), are:

- Storage engine: used to store data and to interface the DBMS with a file system at the operating system.
- Metadata catalogue: Database dictionary, functions as a repository for all database objects/elements.

- Database access language: application programming interface (API) provided by the DBMS to access data. Structured query language (SQL) is an example of a database accessing language.
- Optimization engine: used to parse database access language requests and turn them into actionable commands for accessing and modifying data.
- Query processor: means for running the query and returning the results.
- Lock manager: manages concurrent access to the same data. Required to ensure multiple users aren't trying to modify the same data simultaneously.
- Log manager: ensures that log records are made efficiently and accurately. Logs are records of changes made.
- Data utilities: set of utilities for easier management and control of database activities.

1.3.3 Different database models and management systems

There are several types of DBMS with some being more constantly used than users depending on adaptability and context. A highly used DBMS and the most popular type of DBMS in the market is relational DBMS or RDBMS. MySQL, Oracle, and Microsoft SQL Server database are examples of RDBMS. Relational databases as the name implies derive from RDBMS. These databases have their data relationships defined in the form of tables, this is, rows in tables with a fixed schema and relationships defined by values in key columns.

Besides relational databases, there are more types of databases with different management systems (*Types of Database Management Systems*, n.d.).

Object-oriented databases are included in relational databases but with differences in data storage. Data is stored in the form of objects with attributes and methods that define how to use the stored data.

In hierarchical databases, with hierarchical DBMS, data is stored in a parent-children relationship node. In these databases, besides actual data records, databases also contain records about the relation between elements.

Network databases are hierarchical databases, but the difference is that a node can have relationships with several entities where in hierarchical databases a node can only have a single parent.

NoSQL databases are databases that do not use SQL as their accessing language and have 5 major categories: Column, Document, Graph, Key-value and Object. Examples of NoSQL are CosmosDB, MongoDB, CouchDB, ArangoDB and more.

NewSQL DBMS offer the same performance and scalability as NoSQL DBMS, but with added isolation, durability and consistency. This database model also features in-memory capability and the ability to be deployed in the cloud.

Finally, cloud databases are simply built in and accessed through the cloud and it is managed by a user organization or service provided by the database vendor.

1.3.4 DBMS architecture and types of DBMS architecture

DBMS architecture is, as defined by many, the structure of the database system and how the users connect to it. Architecture is important to consider because it affects the performance, implementation, design and maintenance of the DBMS.

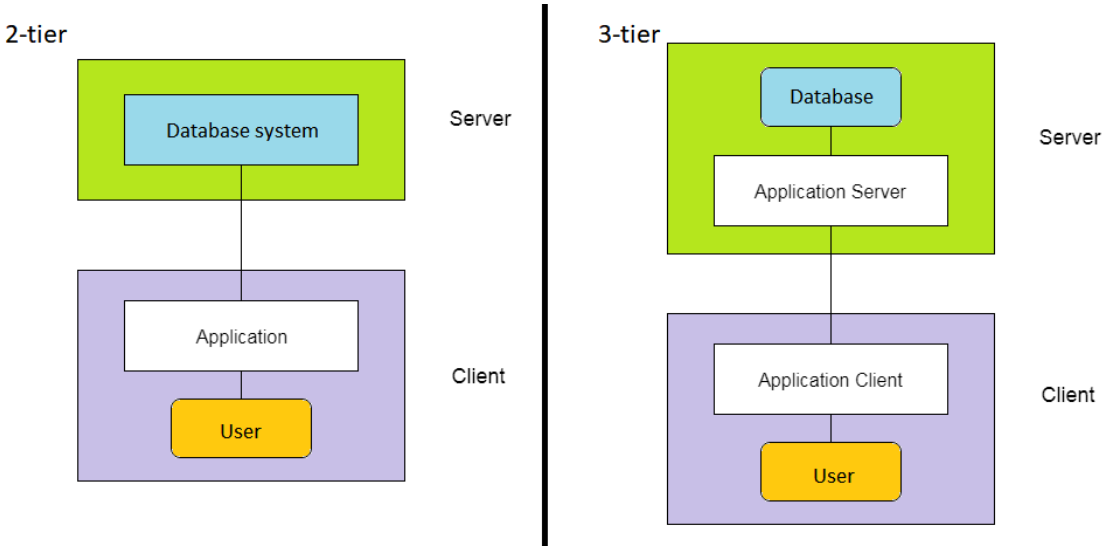

Figure 17: Two-tier (left) and multi-tier (right) DBMS architectures adapted from (*DBMS Architecture - Javatpoint*, n.d.).

DBMS architectures vary according to how users connect to the database servers to carry out their requests. These architectures can be categorized as single-tier, two-tier, or multi-tier (3-tier) (*DBMS Architecture - Detailed Explanation - InterviewBit*, n.d.):

- Single-tier architectures (1-tier): simplest DBMS architecture. Client, server, and database are all on the same machine. There is a severe lack of security in this architecture, and this is only recommended when creating a local application.

- Two-tier architecture (2-tier): client-side applications establish a connection with the server-side to communicate with the database and this connection is possible through APIs. In this cases, the DBMS is not exposed to end-users. Figure 17 presents a simple schematic of 2-tier DBMS.

- Multi-tier (3-tier): most frequently used DBMS architectures. There is an added layer besides the Database layer and Presentation layer, which is the application layer. This application layer helps reduce the server's query processing. Figure 17 presents a simple schematic of multi-tier (right) and 2-tier(left) DBMS.

1.4 Web development: front-end vs back-end

Web development is the process in which the main goal is to create a website and applications for either the internet or private networks. This process can result in simple static pages or in complex web applications and web services.

When entering the world of web development, there is almost immediately a distinction between 2 terms: front-end and back-end. Front-end and back-end are 2 different types of web development with different purposes and tools.

In simple terms, front-end development consists on building the website with the user in mind, in other words, it is a style of computer programming that focuses on creating features and elements that will be seen by the user. Front-end development is the area where the developer must worry about website-user interaction by making sure all the features presented to the user function properly. Front end developers work with

languages like Hypertext Markup Language (HTML), Cascading Style Sheets (CSS), Javascript (JS) and with some front-end frameworks such as AngularJS, ReactJS and JQuery.

Back-end development focuses on the side of the website users can't see. Users do not directly interact with the back end of a website, they will indirectly interact with elements these developers work on through a front-end application (*Front End vs Back End Development: What's the Difference? - Kenzie Academy*, n.d.). This important part of website development is what makes various sites dynamic and interactive and that ensures it all works. It is in the back-end that databases and all its many features (from the storage engine to the log manager already described before) exist. In sum, back end development deals with storing and arranging data while also ensuring the front end is functioning well. Back-end developers work in languages like PHP, C++, Java, Python and Javascript.

1.4.1 Hypertext markup language

Hypertext markup language (HTML) is the basic scripting language used by web browsers to render pages on the world wide web (*HyperText Markup Language – HTML Definition*, n.d.). In simple terms, it is a series of short codes typed into a text-file, and these short codes are called tags. Tags are the basis of the Hypertext markup language and browsers read these file types and translate them to visible and readable forms to the user.

1.4.2 Cascading style sheets

Cascading style sheets or simply CSS is, as the name implies, a style sheet language used modify and describe the elements presented in either HTML or XML documents. It styles all the elements and describes how they should be rendered on screen. CSS is among the core languages of the open web and is standardized across Web browsers according to the World Wide Web Consortium (W3C) specifications (*CSS: Cascading Style Sheets | MDN*, n.d.).

1.4.3 Javascript

JS is a programming language that is one of the most important core language technology of the world wide web. According to a study regarding user statistics, 98% of websites use JavaScript on the client side for webpage behaviour (*Usage Statistics of JavaScript as Client-Side Programming Language on Websites, August 2022*, n.d.).

JavaScript is a complex and high-level interpreted language that conforms to the ECMAScript standard. It is capable of many functionalities, namely dynamic typing and prototype-based object-orientation. It has various APIs for working with text, dates, regular expressions, standard data structures, and the Document Object Model (DOM).

1.4.4 Python

Python is another high-level, interpreted and general-purpose programming language. It is capable of many programming paradigms such as structured, object-oriented and functional programming. This programming language had a first release in 1991 by Guido van Rossum (python 0.9.0).

Python is often called by many programmers one of the easiest languages to learn and that is because of its core philosophy points (from the document The Zen of Python (PEP 20)):

- Beautiful is better than ugly
- Explicit is better than implicit
- Simple is better than complex
- Complex is better than complicated
- Readability counts

Python was designed around extensibility via modules and that is why it is extremely versatile and works in various machines.

1.5 Examples of other polyglutamine disease databases

There are already some existent databases that offer various type of information regarding polyQ diseases. Mouse polyQ database is a public database comprising the polyQ mouse

models, phenotypes and therapeutic interventions tested *in vivo*. This database presents to a certain extent basic information about the mouse models, such as their genetic design, and detailed descriptions of the disease phenotypes and information regarding therapeutic approaches evaluated in polyQ mice, including information about drugs and the phenotypes that were evaluated as therapeutic outcomes (Szlachcic et al., 2015).

Another database that has been developed was PolyQ which, at the date of 14th September 2022, was not available. The authors present the database as a bioinformatic tool with all human sequences from the NCBI non-redundant (NR) database that contained at least seven consecutive glutamine residues (Robertson et al., 2011). It could later be interrogated such that the sequence context of polyQ repeats in disease and non-disease associated proteins could be investigated (Robertson et al., 2011).

Many other databases exist that are not exclusive for polyQ diseases, but they offer various types of information regarding these and other documented human diseases. ORPHANET is a database for rare diseases, a group in which polyQ disorders are included. It includes many services but the most used perhaps are the inventory and encyclopaedia of rare diseases which gives information on the signs and symptoms, classifications, genes, emergency guidelines, sources/procedures and more. The database can be accessed via https://www.orpha.net/.

Online Mendelian Inheritance in Man or OMIM is another database and compendium of human genes and genetic phenotypes which focuses on the relationship between phenotype and genotype. OMIM covers many disease topics including overall descriptions, clinical features, inheritance, mapping, pathogenesis, molecular and population genetics and animal models. OMIM can be freely accessed via https://www.omim.org/.

CHAPTER 2 – OBJECTIVES

2. OBJECTIVES

2.1 Data is scattered on the internet and literature.

Information about polyQ diseases, can be found in articles or journals that describe research studies, scientific review articles and reports on clinical trials. Original articles offer information on the disease of interest as an introductory step to the experiments and results later shown, while review papers may focus on and deepen a single broad topic such as epidemiology, pathogenic mechanisms and therapies.

After these articles, information on polyQ diseases is vastly found in books written by specialists of the area, that elucidate on various subjects and help people better understand disorders in this group.

Finally, there are also databases: databases that have saved data on every disease known to date, including glutamine-expansion disorders, but that information is usually brief; rare diseases databases such as Orphanet (mentioned before), that also include abbreviated data on polyQ diseases; and databases made to store only protein and gene data are also very present nowadays and provide vast and updated information (Uniprot and NCBI gene, respectively).

2.2 PolyQ database objectives

Considering what was mentioned above, the main objective of the current project was to design and establish a novel database, the PolyQ Database, that would collect and centralize information on pre-defined topics for all 9 polyQ diseases. The project aimed also at to organising that information in a concise manner, with original images and schemes. The final goal was to ultimately make the database available online, with free access to anyone, therefore creating a new online resource with extended information about polyQ disorders.

CHAPTER 3 - METHODOLODY

3. METHODOLOGY

3.1 Deciding the topics included in the database

Initially, it was decided that common topics across every polyQ disease would be chosen to be included when describing these diseases in the database website. The discussion about which topics to be chosen had several steps and matters of concern. Hoping for a simple yet complete database about polyQ diseases, it was necessary to focus on several disease aspects that interested the visitors, which could be students, scientists and the general public. Overall, one of the most searched topics when looking for polyQ diseases is how these diseases manifest clinically and another one is how one acquires the disease which in these cases is directly linked to genetics and inheritance. After the 2 most searched and commonly presented topics, other important information is the mutation itself, the information on the gene and protein along with data on clinical trials, review articles and other databases information.

Further discussion led to the hypothesis that the database had to target all types of interested audiences, therefore, it had to present simple information to introduce each disease but include also the fundamentals of the mechanisms that have been suggested to be involved in each disease along with the necessary informative tools. Figure 18 presents a flowchart of the structure of all the topics to be included for each disease, which was used to organise the discussed ideas and conclusions (topics presented were not exactly represented in the website as the flowchart shows).

3.2 Searching and organizing information

Information on polyQ diseases focusing the discussed topics was found on various online sources along with theme-related books. Most of the information found on the different topics was assessed by an initial google search with the name of the disease followed by the topic. This same keyword search was performed in scientific databases. Importantly, when searching for general disease information to be used in the introduction, for instance, it was not necessary to add any topic as a search keyword, only the disease name sufficed.

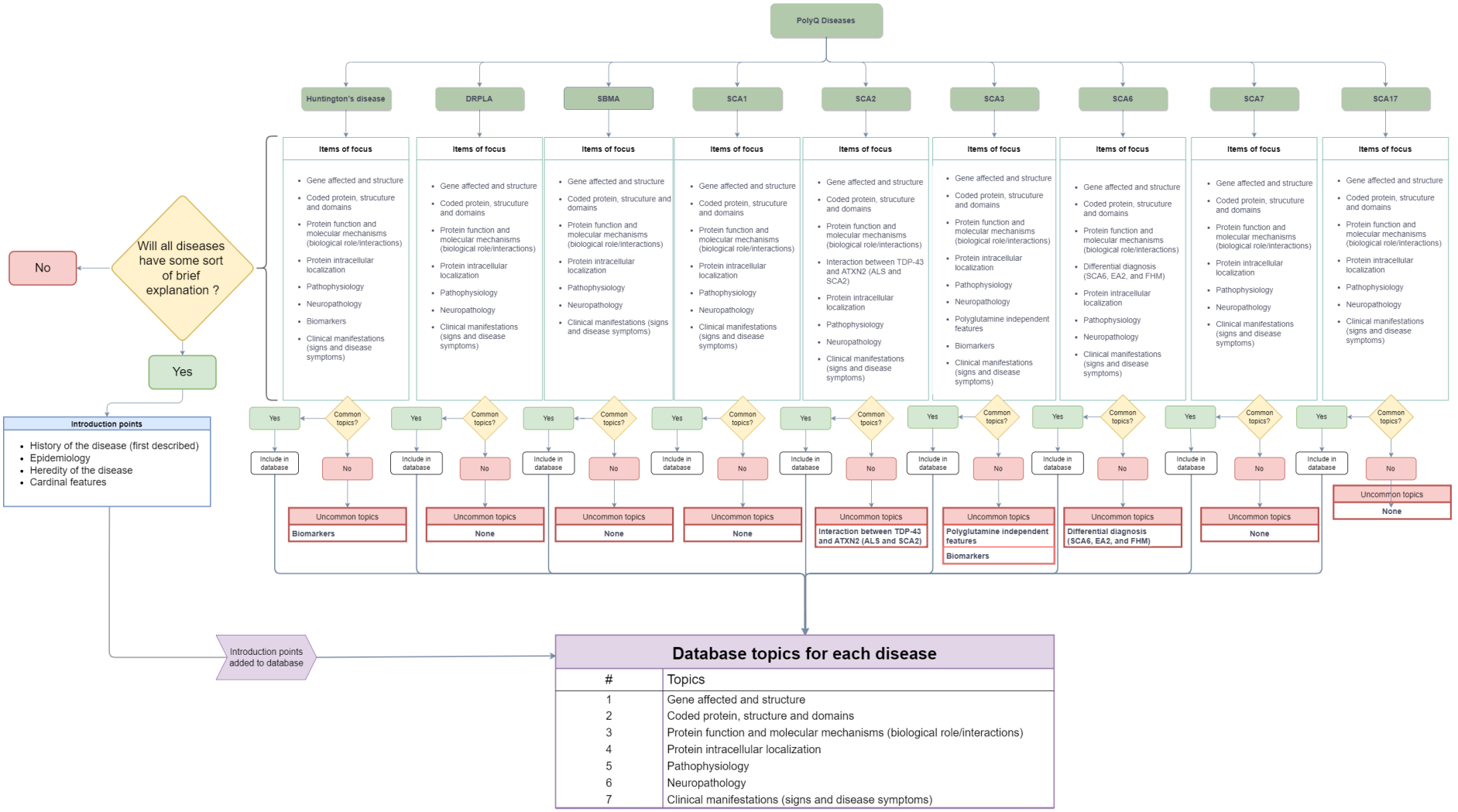

Figure 18: Flowchart used to organise ideas and determine the topics to be included in the database

Most of the information was obtained from scientific articles which were found in scientific article databases such as the National Center for Biotechnology Information (NCBI): Pubmed and Pubmed Central, ResearchGate, ScienceDirect and Wiley Online Library. Many renowned journals websites such as Nature, Cell, Journal of Biological Chemistry were also used. All these online tools provided research articles, book chapters and reviews regarding all topics of every disease. To further complete the information about the genes and proteins associated with the 9 diseases, Ensembl, NCBI: Gene and UniProt were used. Ensembl gave important gene information, namely size, coding regions, number of exons and more. NCBI provided gene size and variants. UniProt offered most of the information regarding the affected proteins: size, domains, expression pattern and functions.

When organizing information, all the references were stored using Mendeley for both future referencing and extracting information. Besides Mendeley, all 9 diseases had an original manuscript made in Microsoft Word with all the information properly segmented and structured with the previously mentioned topic layout.

3.3 Original images

All images presented in the website are original and made with the help of software tools.

The first software used was Biorender, which is an online browser tool with many pre-made icons and templates from all the biology scientific fields. Biorender is easy to use and offers intuitive visuals and is accessible to everyone. When designing simple linear representations of gene and proteins, Biorender was used as a fast and reliable method.

For more complex images and schemes, namely the ones describing pathophysiological mechanisms, Adobe Illustrator was used. After learning most of the software functionalities, Illustrator proved to be versatile and enabled the creation of complete schematic representations that accompanied the information in each of the disease manuscripts.

3.4 Learning HTML, CSS and JS. Front end tools

Instead of relying on a content management system such as WordPress and Squarespace, it was decided to create a website from scratch with HTML, CSS and JS for the front-end

part. These tools seemed enough for what was initially purposed and both HTML and CSS are simple and straightforward languages for learning and implementing. JS, on the other hand, required more time to understand all the basic functionalities and implement them. Also, a JS framework called jQuery was used to facilitate certain mechanisms within the website. This framework is simple and has great documentation to follow in order to make it work in production applications. Besides frameworks, some JS scripts with MIT licenses made by other authors were also used namely the animate-on-scroll script which is present in some website pages.

Most of the learning of these 3 languages was made through online sources and tutorials and any punctual doubts were often resolved in programming forums such as Stackoverflow. Also, there are extended documentations on all these 3 languages from various trusted sources that were visited and used. Even more, browsers themselves have a developer tab which contains various information on the tools used to make the browser work and these were very important in code implementation.

3.5 Python and Django library. Back-end tools

For future proofing, dynamic pages were created within the website. Python was used to create a simple yet sufficient website back-end. From all the extensive Python libraries and packages, Django seemed the best and easiest to implement with all the needed functionalities in mind. This library is a high-level Python web framework that encourages rapid development and clean, pragmatic design and it takes care of much of the hassle of web development (*The Web Framework for Perfectionists with Deadlines | Django*, n.d.).

There was some previous understanding of the Python language, but Django was something new. Learning the basics of Django was simple and straightforward once again thanks to the extensive documentation provided along with forums and other online teaching and informative sources.

As the aim was to develop a website that presented clear and updated information on polyQ diseases, we needed the information-providing pages to be dynamic, in order to easily change, correct and update information, as science evolves every day.

3.6 Other tools: Atom

All of the coding in HTML, CSS, JS and Python was done in Atom (figure 19). Atom is a free and open-source text and source code editor for macOS, Linux, and Windows. It is a desktop application built using web technologies and developed by GitHub. It was developed with HTML, JavaScript, CSS, and Node.js integration and developers call it a "hackable text editor" because it is fully customizable in HTML, CSS, and Javascript (*Getting Started with Atom | Codecademy*, n.d.).

Atom, as many other code editors, makes web developing easy because of its many features and downloadable packages that increase productivity and efficiency, live-server is one example of a downloadable package that enables the developer to see the changes made in real-time. Overall, Atom worked wonders and had no issues throughout the entirety of the front-end and back-end coding.

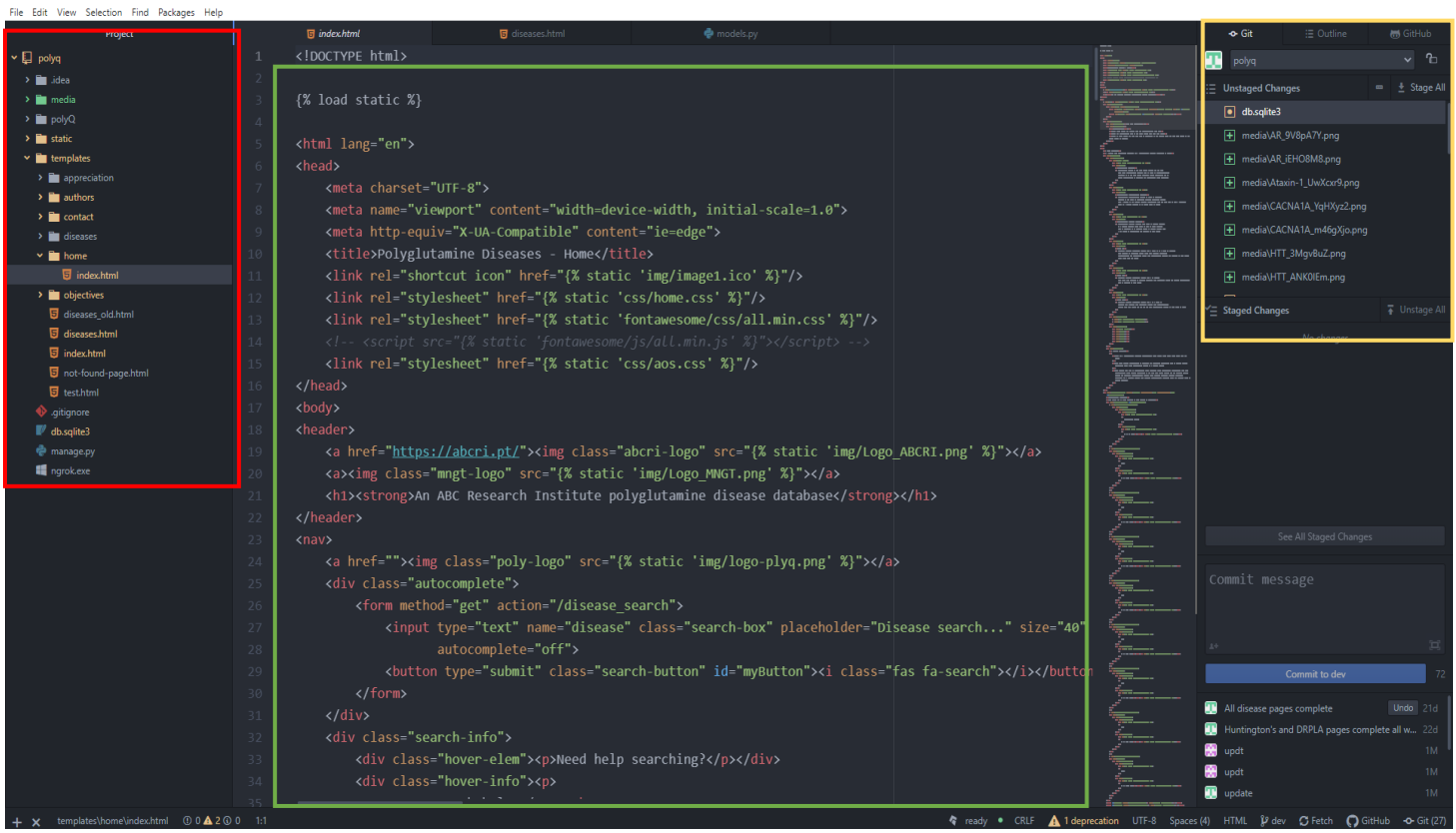

Figure 19: Atom interface. Folder and file browser in the red rectangle, code editor represented in green rectangle and inside the yellow rectangle the GIT managing tools can be found.

3.7 Website creation

The website structure comprised several static pages besides the dynamic disease pages. Besides the dynamic disease pages, all static pages gave pieces of unique information which can be browsed from one to another with options.

All pages were made regarding responsiveness, so every single page has versions for all devices and every single page besides the landing page has a footer section with hyperlinks to all static pages (disease pages excluded).

3.7.1 Landing page

The first static page created was a Landing page where there were hyperlinks to all the other static pages created. Also, it gave some introductory information and redirect button to the home page. To note that the Landing page and the Lome page have different purpose and goals and are not the same.

For this page, HTML, CSS and JS were used along with an original website logo, the research group logos and university logo.

3.7.2 Authors page

Another important static page created was the Authors page. In this page, information on the authors of the PolyQ database project is given with the purpose of introducing the website creators to the people visiting. Also, links to any of the authors social media and other are included.

HTML and CSS were sufficient to create this simple static page.

3.7.3 Objectives page

The Objectives page elucidates the main objectives of the PolyQ database. Questions such as to why it was created or why it is different are answered in this page.

HTML, CSS and JS (for animated visual effects) made possible the creation of the Objectives page.

3.7.4 Contact Page

For visitors to contact the website creators, a Contact page was created. This page only displays emails for contact which visitors can use to communicate with the authors or notify them about website errors or other matters that may be of interest. It also shows the university location through an embed map from Google Maps.

It was decided against a contact form, but maybe in the future that form will be there for automatic messages and to make all communication easier.

This page was made by simply using HTML, CSS and Google Maps Embed API URL.

3.7.5 Appreciation page

As a token of acknowledgement and gratitude, an Appreciation page was created, where all the most important sources and tools are mentioned with the respective hyperlinks. Overall creation of the website, with all its functionalities, was made easier because of all the existing information databases and useful tools, therefore deserving mentioning and praise.

This page was made with the usage of HTML, CSS and JS.

3.7.6 Home page

One of the most important pages is the Home page because it “sets the stage” for the rest of the website. In this page a search bar with autocomplete features was created where visitors can type the polyQ disease of interest. Besides the search bar, there are informative sections on what to expect from the Disease pages. There are also direct hyperlinks to all the available Disease pages as an alternative to the search bar.

The home page was made using HTML, CSS, JS for the front-end and the necessary back-end was made with Python (Django library).

3.7.7 Disease pages

As mentioned above, Disease pages were made to be dynamic for easier information management. Being dynamic means that only one page was enough for all 9 presented diseases because depending on what diseases are searched or clicked the information that will load will come from a database with the help of a back-end system that will load the correct information for each disease.

On every disease page the informative topics were structured as followed: introduction, first description of the disease, epidemiology, causative gene, codified protein, pathophysiology and clinical manifestations. On computer resolutions, besides disease information a topic navigation bar to the left and a reference bar shortcut to the right were included. This is helpful if visitors want to go back to the Home page or choose another disease. Buttons with hyperlinks to those locations are present in the top navigation bar. In case of mobile devices, mobile phones do not have the topic navigation bar and reference shortcut bar and references are shown after all information topics. On most tablets, the topics navigation was made to be visible and functional.

Back-end and database for the Disease pages were made with Django (Python web framework package).

3.7.8 Choosing domain and website deployment

Website domain was chosen to be polyq.pt and was deployed in a virtual private server or VPS with Ubuntu OS. VPS access was done using a secure shell (SSH) protocol. Inside the server, all the packages and files used for the website were cloned to it using git tools, this includes all the front-end and back-end files. Next, virtual host files and apache were configured inside the virtual machine/server with all the required permissions, paths and domains. With all this configuration properly done and tested, the website was successfully deployed to the pretended URL.

3.7.9 Database

The first database, an object-oriented database, was created in Django. In this database, all the information and images across every topic were saved in the form of objects and organized in a working hierarchy.

While deployed to the VPS, it was decided to import the contents of this database to a Microsoft SQL server for better productivity. Docker was installed and used inside the VPS in order to install a Microsoft SQL server. Docker is an open platform for developing, shipping, and running applications. After Microsoft SQL server was installed and ready for production, all data stored in the previous Django database was imported to the new database using Python packages that enabled exporting the contents of the previous OOP-database to xls formats that could be imported in the new database.

CHAPTER 4 - RESULTS

4. RESULTS

4.1 Website structure

The final structure of the PolyQ Database (accessible at polyq.pt) comprises an initial Landing page which can redirect the visitor to a Contact page, an Appreciation page, an Authors page and the Home page. While in the Home page, visitors can either search for the disease of interest or simply scroll down and select one of the available diseases by tags, this way being redirected for the selected Disease page (figure 20).

4.2 Online interface of all pages besides Disease pages

The PolyQ Database interface was designed to be easy to browse with all the options at the disposal of the visitor. This section represents the static pages interface as shown in computer devices with 1920x1080 resolution as the best example, but all website pages are responsive in any other devices and resolutions, as mentioned in Chapter 3.

4.2.1 Landing page interface

The first page that loads when visiting PolyQ Database is the landing page (figure 21).

On top of the page inside the green rectangle, there are buttons for each of the pages the visitor decides to click on. In the red rectangle, there is the main button that simply redirects to the home page.

4.2.2 Objectives page

One of the supplementary pages is the Objectives page. This page answers 3 questions that reveal the reason for the creation of the PolyQ Database. The display is shown in figure 22.

On top of the page there is the header that is present in all pages besides the Landing page. Right under it, marked with the green rectangle, there are 2 buttons to redirect to either

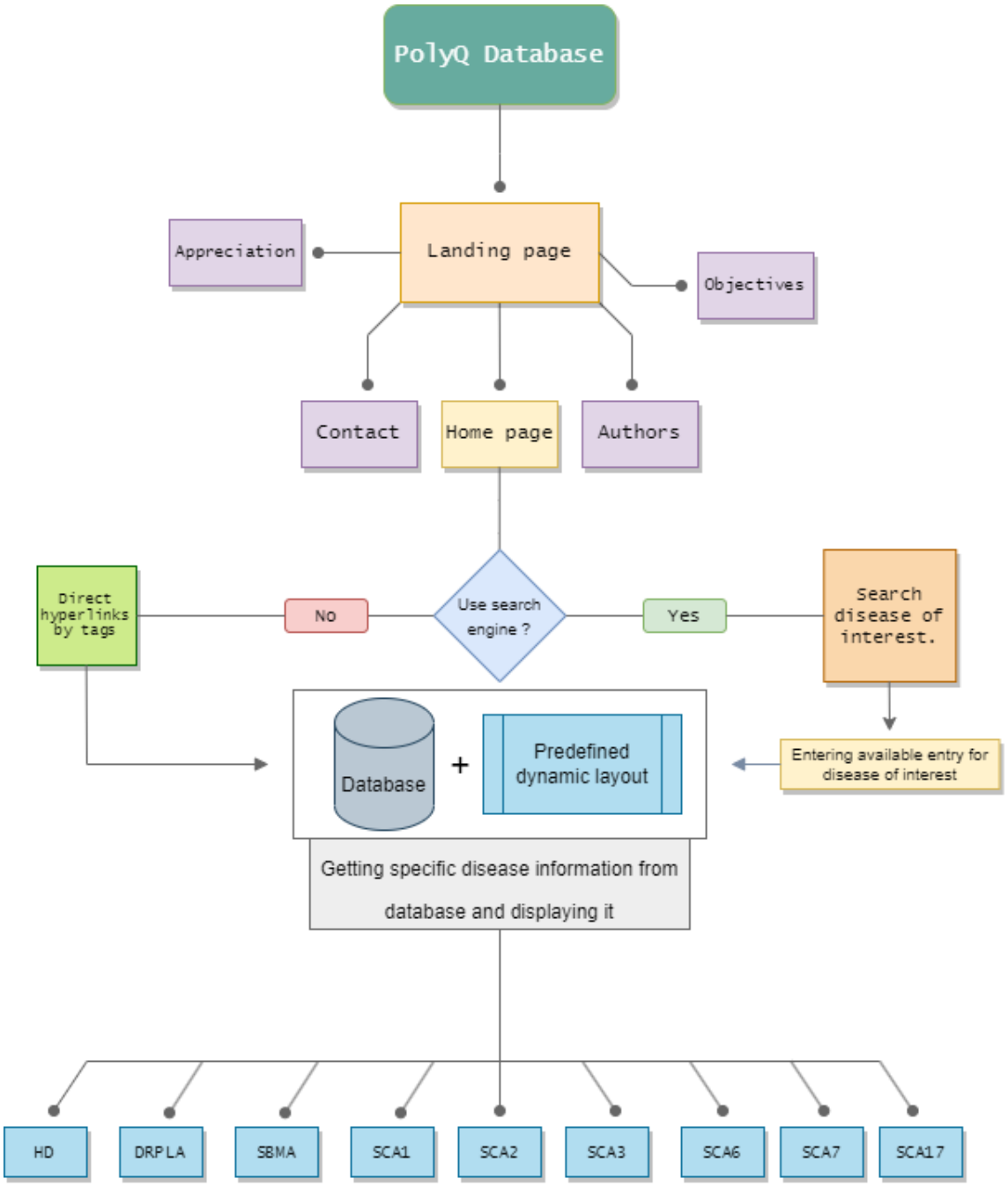

Figure 20: PolyQ Database website structure

the Home page or the Landing page. In orange rectangles there are the questions meant to be answered and that can be expanded to show the answer (figure 23). The red rectangles surround the answers for each question, and these can be collapsed and go back to the original layout shown in figure 22.

Figure 21: Landing page interface

4.2.3 Contact page

For contacting purposes, there is a Contact page with available email addresses and links to social media pages, to check and message. The appearance of the Contact page is shown in figure 24. The available contact emails are shown inside the blue rectangle.

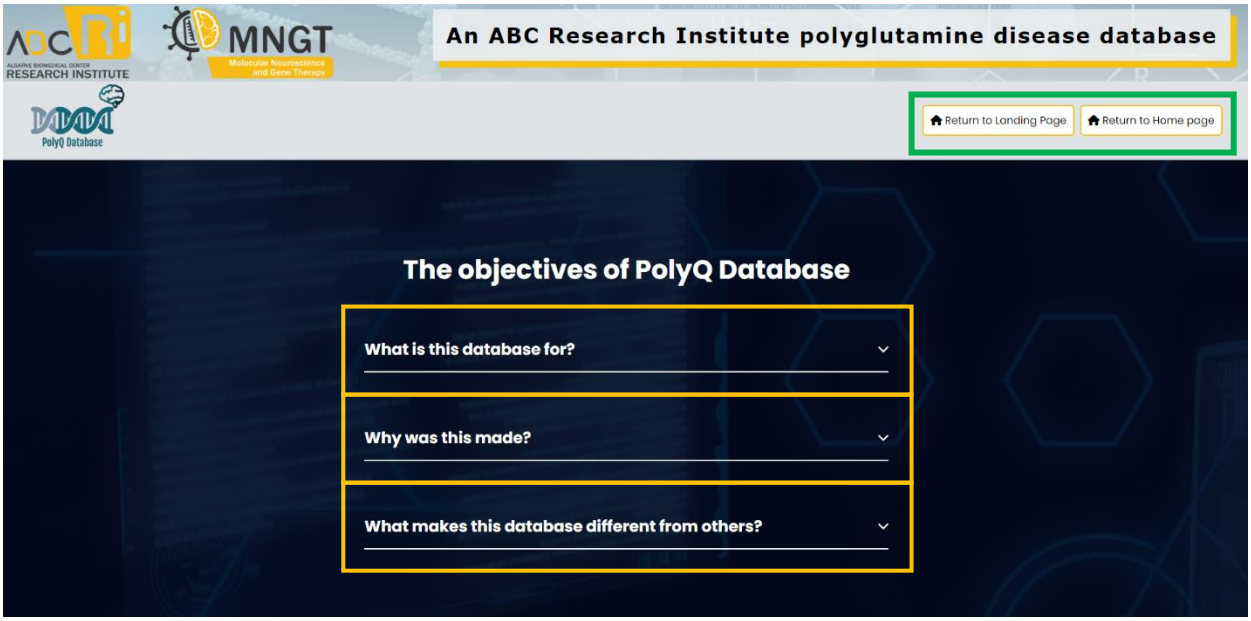

Figure 22: Objectives page interface

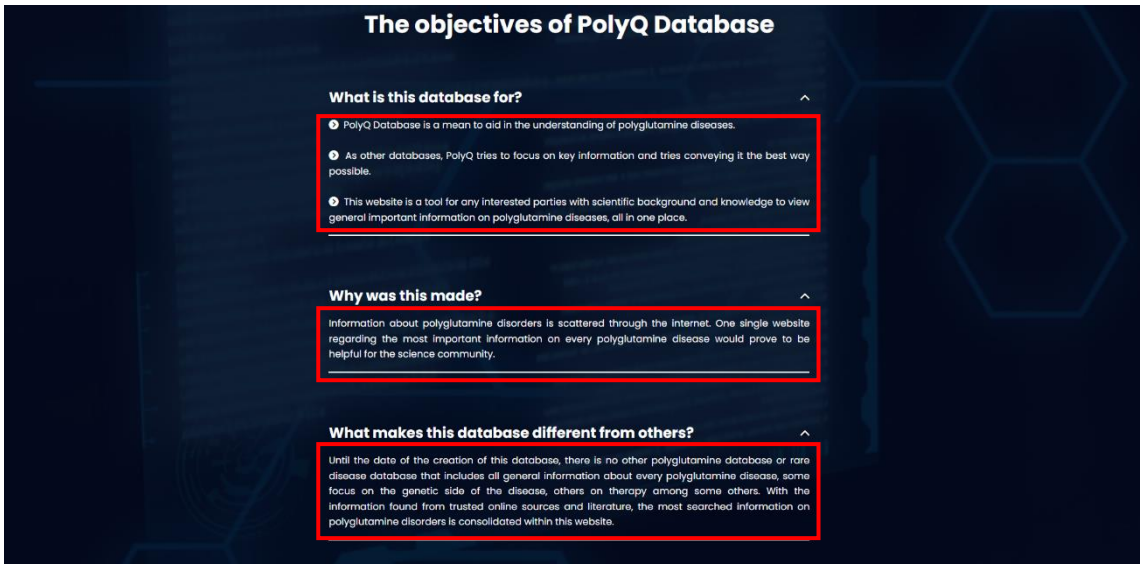

Figure 23: Objectives page with questions expanded

Socials that can be messaged are shown inside the green rectangle and an embedded map of the location of the Campus de Gambelas of the University of Algarve is shown inside a yellow rectangle.

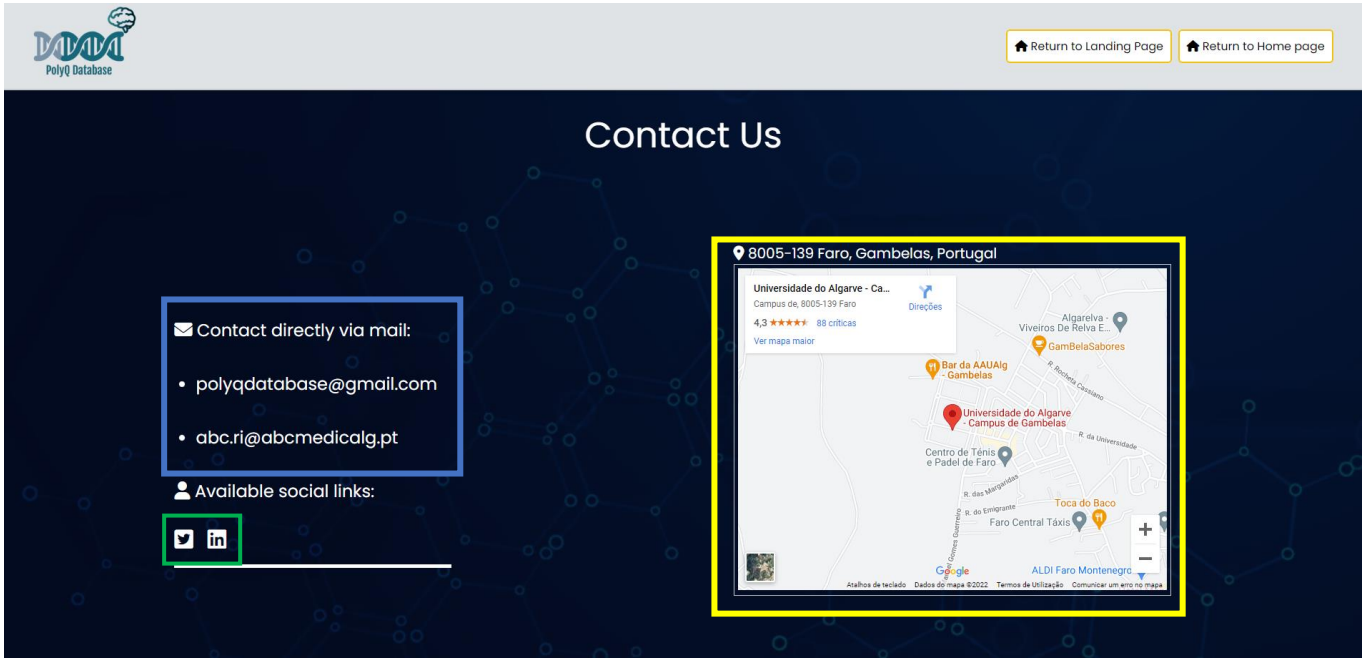

Figure 24: Contact page interface

4.2.4 Authors page

An Authors page was created to inform the website visitors of the main people involved with this project. Three authors were mainly responsible for this website creation and there are small pieces of information and social links for every author. The page is displayed in figure 25. Each green rectangle presents an author with the respective image and briefly described professional backgrounds. Surrounded by a blue rectangle are the socials that can be accessed directly for each author.

4.2.5 Appreciation page

For all the tools and scientific sources and databases that helped build the website, there is an Appreciation page that presents all of these with the respective links and messages of gratitude. Figure 26 shows part of the display of this page.

4.2.6 Home page

It is the Home page that links the user to the disease information pages. Besides having the search engine and direct links to the diseases, it has other informative elements, as

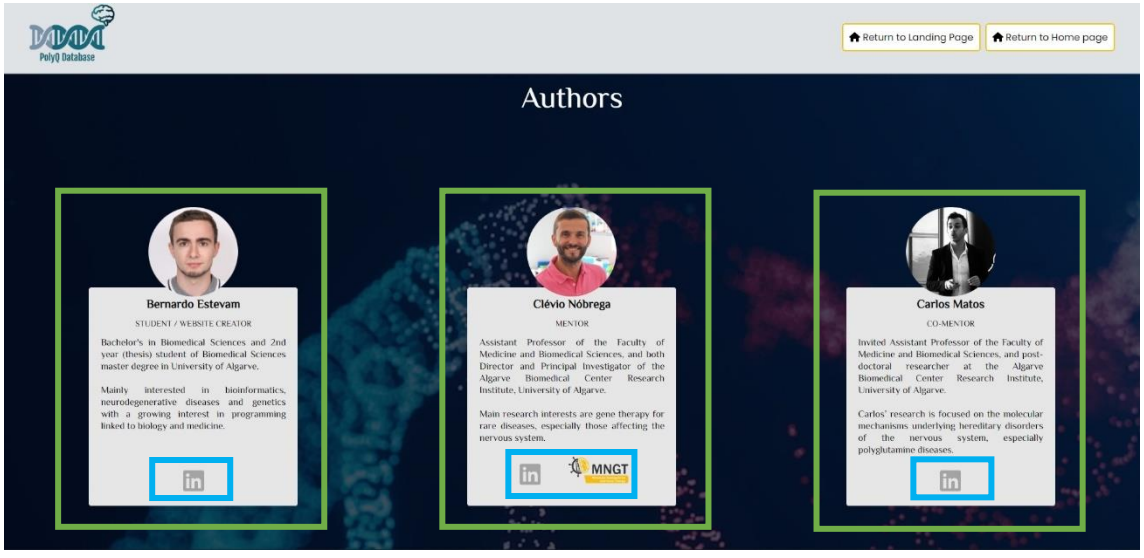

Figure 25: Authors page interface

mentioned before in Chapter 3. The top interface and tags of the Home page are shown in figures 27 and 28, respectively.

Figure 27 represents several important elements. Firstly, in the green rectangle there is the search bar where visitors can start typing and the suggestions of the autocomplete function appear. If there is a match with the search criteria the page is redirected to the desired Disease page. The search criteria are briefly explained with a tool tip marked with the red rectangle. Users only need to move their cursors into the tool tip and a box with search help is displayed. Slightly below, there is an element marked with a yellow rectangle. This is an image slider that gives small pieces of information on polyQ diseases. As the Home page is browsed the image slider can be manually clicked to change images or it automatically swaps images overtime and it presents small introductory pieces of information regarding the PolyQ diseases group. Lastly, in a blue rectangle, there are 3 bullet points that elucidate the visitor on what to expect from the website.

Figure 26: Appreciation page interface

Figure 27: Top of the interface of the home page

The Disease pages direct links are represented in figure 28 as shown in the website. By clicking any of the boxes (surrounded by green rectangles) the user is redirected to the corresponding Disease page. Inside the red rectangle is the navigation arrows for choosing other tags that may be off screen.

4.3 Disease pages

4.3.1 Topics described in the Disease pages

The topic structure of each Disease page, summarized below, was envisioned after considering all the information that was selected to be displayed and the way it would be presented:

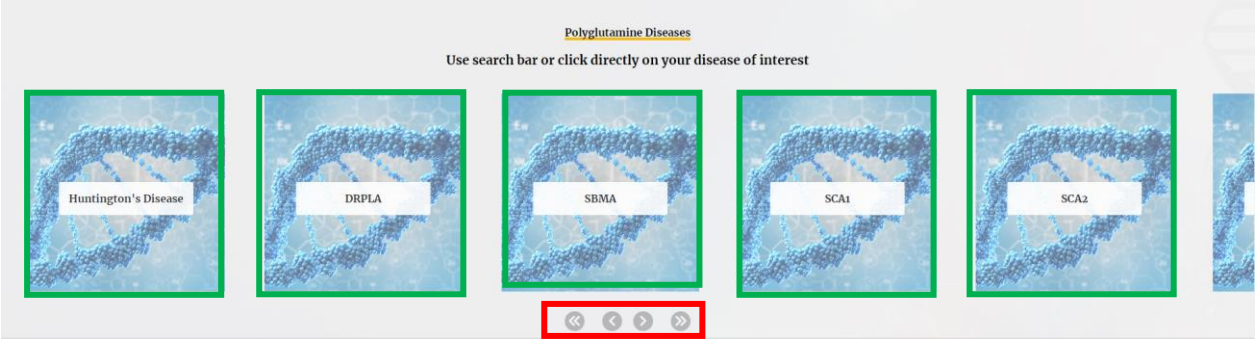

Figure 28: Home page clickable tags with direct links to chosen Disease page

- Introduction: brief introductory information on the disease (gene affected and clinical age of onset in some Disease pages are briefly presented in the introduction but further explained in other topics) and inheritance features;
- First described: first description of each disease with authors referenced. This topic presents when and where cases of the disease in question were first described;
- Epidemiology: most important epidemiological features (around the globe) of every disease. An original map with information from various research articles is presented, conveying the information as a simple and illustrative supplement;
- Causative gene: information on the gene affected in each disease along with simple gene structure representation and information. Information such as the name of the gene, where it is located, how is the gene mutated and where exactly, and different degrees of mutation are all presented in this topic;
- Codified protein: domain structure information along with normal protein functions and roles and where the protein is localized intracellularly are conveyed here. Important to note that inside this topic the information is organized in further subtopics to split and display protein data in a more simplistic manner. Structure and domains, protein functions and biological role and protein intracellular localization are subtopics;
- Pathophysiology: information of the most important pathophysiological mechanisms that lead to disease progression and worsening. As a supplement to the text, an image is displayed that translates the textual information to a schematic;
- Clinical manifestations: clinical topic explores the neuropathological findings in patients affected by each particular disease and the symptoms they show along disease progression. As is the case of the protein topic segment, this final topic divides the information into subtopics, which are neuropathology and signs and symptoms.

4.3.2 Disease pages interface tools

For better understanding of the functionalities that every Disease page has and how they are supposed to work, this subsection shows the general interface and tools present across each one of them.

When loading any Disease page, there are 3 initial functionalities, as observed in figure 29. Inside the green rectangle there are navigation options. Users can either choose any of the other 8 diseases or go back to the Home page. Inside the red rectangle, there is a useful page navigation tool that users can interact with and click to move to the topic of interest. Finally, in the blue rectangle there is a reference shortcut list. Throughout any Disease page references will show up in the text and, by clicking the number of the reference, the reference shortcut list will highlight the selected reference. Figure 30 shows this functionality. Important to note that, in devices with smaller resolutions, this reference shortcut list does not exist, only the references topic on the bottom of the page.

Across the various topics there are images that can be downloaded for better viewing. From the maps to protein structures, all the images are original. There is also a schematic of a cell that represents the intracellular location of every protein associated with each

Figure 29: Top of Huntington's disease page

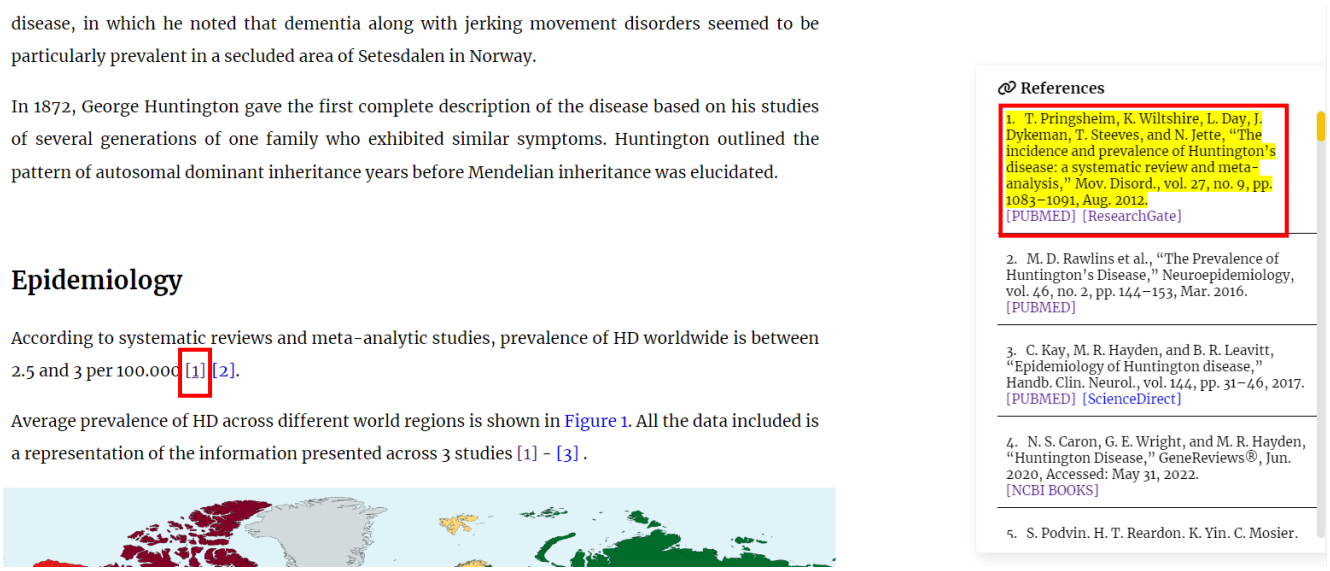

Figure 30: Reference focus list example. By clicking the 1 (small red rectangle), the list highlights the reference in the list on the right.

polyQ disease, as exemplified in figure 31. Depending on the disease, these location points and descriptions vary.

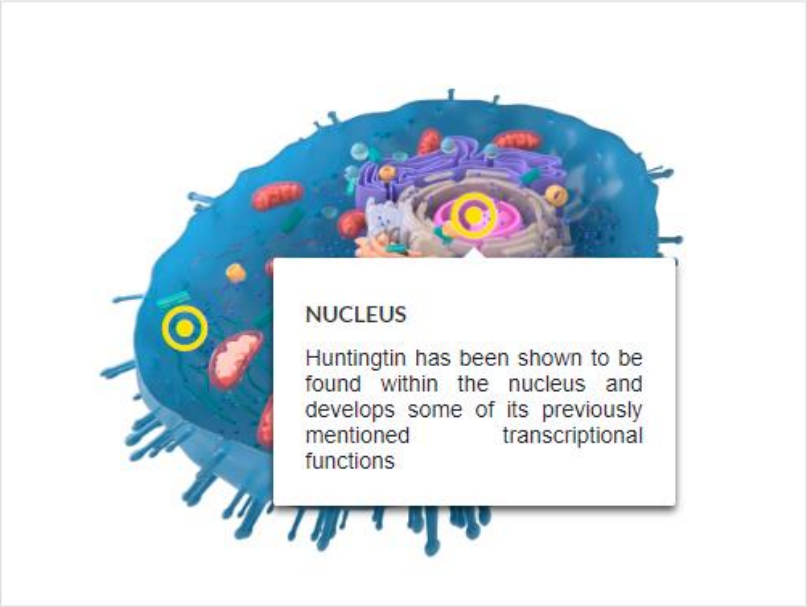

Figure 31: Schematic of a cell present in all Disease pages. Every marker has the name of the location and a description of how the protein may be found there.

4.4 Examples of Disease pages

This segment shows the information collected for each disease, using the pages on Huntington's disease and SCA1 as examples. The text of every topic and the original images included in the PolyQ Database website are reproduced below, in a format adapted to the present document.

4.4.1 Huntington's disease

4.4.1.1 Introduction

Huntington's disease (HD) is an inherited condition that causes progressive degeneration of neurons in the brain. Signs and symptoms usually develop between 35 to 44 years of age and may include uncontrolled movements, loss of intellectual abilities, and various emotional and psychiatric problems. People with HD usually live for about 15 to 20 years after the condition begins. It is caused by changes (mutations) in the *HTT* gene and is inherited in an autosomal dominant manner. This means that having a mutation in only one of the 2 copies of the *HTT* gene is enough to cause the condition. When a person with HD has children, each child has a 50% (1 in 2) chance to inherit the mutated gene and develop the condition.

Treatment is directed to the symptoms present in each person and may include various medications.

4.4.1.2 First described

The first recorded mention of HD was made in 1842, in a letter written by Charles Oscar Waters and published in Practice of Medicine. In 1846, Charles Gorman noted that the disease seemed to occur in particular regions. In 1860, Johan Christian Lund also gave a description of the disease, in which he noted that dementia along with jerking movement disorders seemed to be particularly prevalent in a secluded area of Setesdalen in Norway.

In 1872, George Huntington gave the first complete description of the disease based on his studies of several generations of one family who exhibited similar symptoms.

Huntington outlined the pattern of autosomal dominant inheritance years before Mendelian inheritance was elucidated.

4.4.1.3 Epidemiology

According to systematic reviews and meta-analytic studies, prevalence of HD worldwide is between 2.5 and 3 per 100.000 individuals (Pringsheim et al., 2012; Rawlins et al., 2016).

Average prevalence of HD across different world regions is shown in figure 32. All the data included is a representation of the information presented across 3 studies (Kay et al., 2017; Pringsheim et al., 2012; Rawlins et al., 2016).

4.4.1.4 Causative gene

The *IT15* gene, also known as huntingtin (*HTT)* gene, is affected in HD. There is an increased amount of CAG repeats (CAG expansion) in exon 1 of this gene in HD patients,

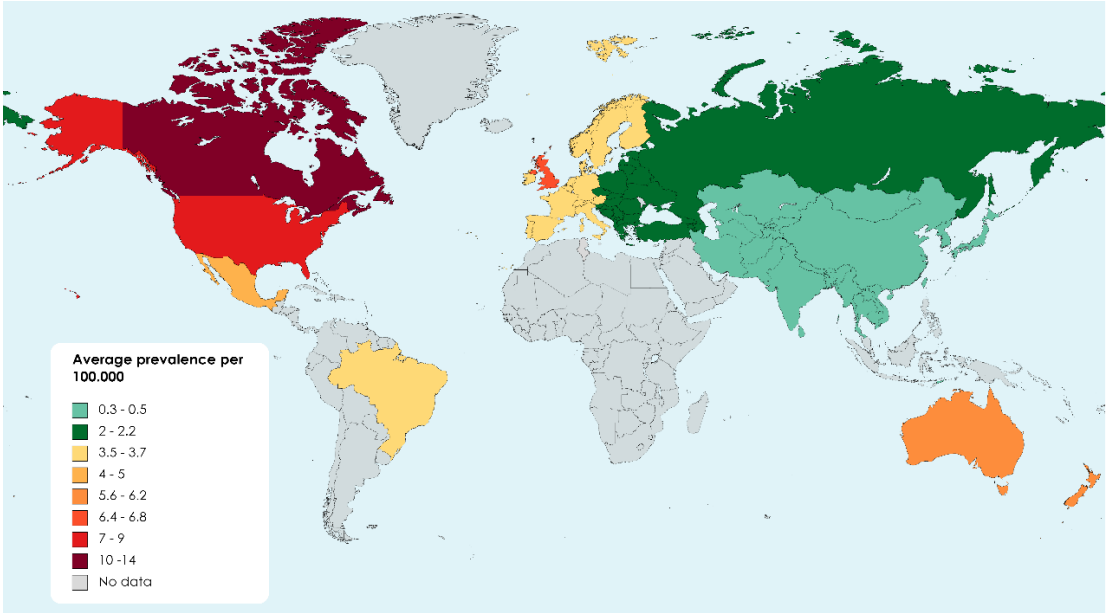

Figure 32: Huntington's disease epidemiologic features.

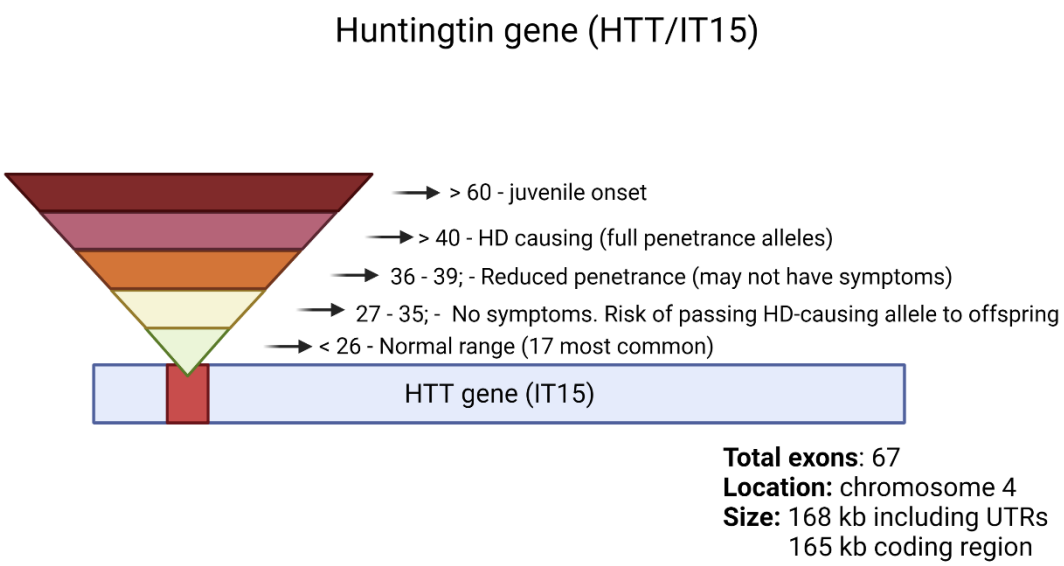

Figure 33: *HTT* gene with ranges of CAG repeat numbers associated with HD, juvenile onset of HD, healthy individuals and other conditions, along with information about the total number of coding exons, chromosome location and size in DNA bases.

usually more than 39 repeats. Numbers of repeats below or way above this threshold have documented implications that are demonstrated in figure 33 (Caron et al., 2020; Podvin et al., 2019; Rego & de Almeida, 2005). The *HTT* gene has a total of 67 exons, is located in the short arm of chromosome 4 (4p16.3, cytogenic band) (figure 34) and codes for the protein huntingtin (Htt).

4.4.1.5 Codified protein: structure, domains, functions and intracellular localization

4.4.1.5.1 Structure and domains

Huntingtin (Htt) is a soluble 3.144 amino acid protein (~350 kDa), found mostly expressed in the central nervous system (CNS) and testes. The abnormal polyglutamine expansion in the N-terminal region of Htt produces significant dysfunction and neural death (Jimenez-Sanchez et al., 2017).

The N-terminal Gln-rich and Pro-rich domain has great conformational flexibility and is likely to exist in a fluctuating equilibrium of alpha-helical, random coil, and extended conformations. A simple Htt protein representation is shown in figure 35, and table 3

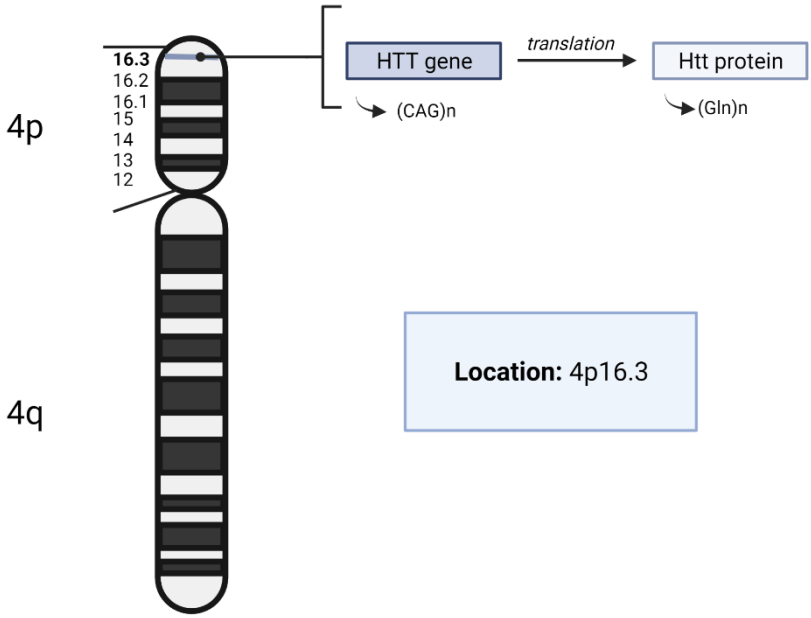

Figure 34: HTT gene chromosomal band localization.

summarizes information on each important domain of Htt (information from Uniprot.org).

4.4.1.5.2 Protein functions and biological role

Htt has various important cellular functions and biological roles:

- Essential for embryonic development, as HTT gene knockout has been shown to be lethal in mice by embryonic day 8.5 (Macdonald et al., 1996);
- Increased cellular death in HTT knockout animals indicates an anti-apoptotic role for Htt, which is supported by the findings that overexpression of full-length wild-type Htt in cultured striatal neurons protects them from apoptosis, whereas depletion of wild-type Htt makes neurons more sensitive to apoptotic stimulus (Leavitt et al., 2006; Rigamonti et al., 2000; Y. Zhang et al., 2003);
- Htt can bind and prevent the formation of the pro-apoptotic Hip1-HIPPI (Hip1 protein interactor) complex (Gervais et al., 2002);
- Htt can directly block the activation of caspase 3 and 9 (Rigamonti et al., 2000, 2001);

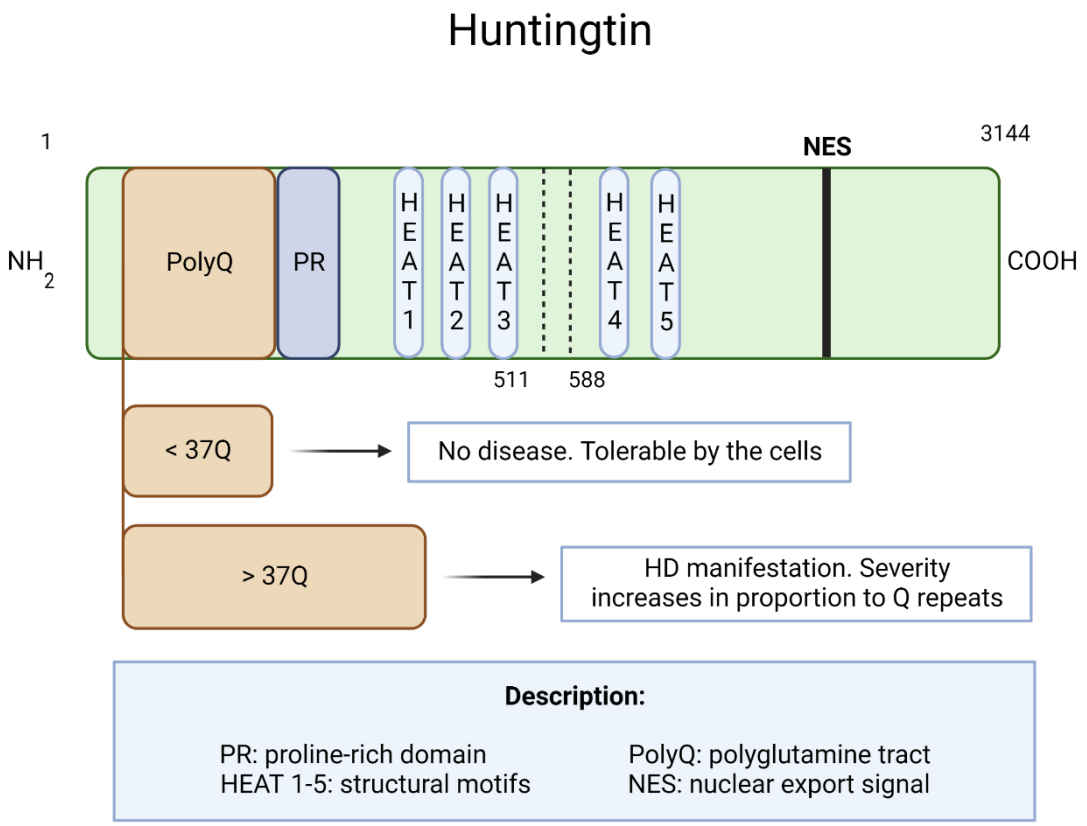

Figure 35: Huntingtin protein structure. Caption in figure.

- Htt regulates the production of the brain-derived neurotrophic factor (BDNF) (Zuccato et al., 2001; Zuccato & Cattaneo, 2009);
- Htt acts as a positive regulator of BDNF transcription by preventing the recruitment of REST/NRSF to its DNA response element (Chiara Zuccato et al., 2003, 2007);
- By interacting with a variety of endocytic/trafficking proteins, such as α-adaptin, HIP1, HIP14, HAP1, HAP40, PACSIN1 and SH3GL3, clathrin, and dynamin, Htt participates in long and short range axonal transport and vesicle trafficking by associating with vesicles and microtubules (DiFiglia et al., 1995; Hoffner et al., 2002), stimulates BDNF vesicular trafficking in neurons (Gauthier et al., 2004; Her & Goldstein, 2008), regulates synaptic activity in neurons and regulates synaptic transmission and plasticity (Smith et al., 2005).

Table 3. Huntingtin most referred domains (repeats)

Description	Position	Length
Repeat 1	204-241	38
Repeat 2	246-283	38
Repeat 3	316-360	45
Repeat 4	802-839	38
Repeat 5	902-940	39

4.4.1.5.3 Protein intracellular localization

Immunofluorescent studies with cells in culture have supported both a cytoplasmic and nuclear localisation of Htt (Hughes & Jones, 2011; Saudou & Humbert, 2016).

Many initial studies have supported Htt as an exclusively cytoplasmic protein (Bhide et al., 1996; Sharp et al., 1995) and others have shown nuclear localization of Htt (Kegel et al., 2002; Tao & Tartakoff, 2001).

Currently, Htt is admitted to be largely cytoplasmic, but it can also be found with the cell nucleus (Saudou & Humbert, 2016).

4.4.1.6 Pathophysiology

HD neuropathological mechanisms are documented and mostly surround proteolytic cleavage of mutant huntingtin (mHtt), misfolding, aggregation, transcriptional problems/dysregulation, mitochondrial dysfunction and abnormal cell clearance of Htt (through the ubiquitin-proteasome system and autophagy).

Mutant Htt proteolysis results in N-terminal fragments, where the polyglutamine tract is located. These fragments are more prone to aggregate and cause neuronal toxicity. It is believed that mHTT aggregation and further cleavage by activated proteolytic caspases are directly involved with cell death (Halliday et al., 1998). Protease inhibitors, or

mutations that may block Htt cleavage have been demonstrated to slow down HD progression in mice (Graham et al., 2006; Wellington et al., 2000).

When it comes to misfolding and aggregation of mHtt, the topic becomes less clear. Inclusions are often seen in the cell nucleus but also form in the cytoplasm, dendrites and axon terminals, though to a lesser extent. Htt fragments aggregate depending on the size of the polyglutamine tract, fragment size, and intracellular protein interactors (Chen et al., 2002; Hackam et al., 1998; Li & Li, 1998; Martindale et al., 1998). Inclusions are a heterogeneous population comprising different forms of mHTT, and their definitive role (whether toxic or protective) has not yet been established (Arrasate & Finkbeiner, 2012).

Transcriptional problems that arise in HD are usually related with altered function/activity of several transcription factors and cofactors that Htt interacts with. mHtt (with expanded polyglutamines) has been reported to bind and change the activity of several transcription factors and other effectors (Dunah et al., 2002; Okazawa, 2003). When mHTT binds these transcriptional factors/cofactors it sequesters them away from the normal DNA elements they interact with. Normal Htt binds cytosolic REST/NRSF and normally prevents its translocation to the nucleus, whereas mHtt interacts less with REST/ NRSF, resulting in the accumulation of this factor in the nucleus (Zuccato et al., 2003, 2007). Mutated Htt also occupies the promoter region of PGC-1α by associating with the CREB–TAF4 complex. This leads to suppression of PGC-1α, which in turn leads to impaired energy metabolism, observed in HD (Trushina & McMurray, 2007). Httt could also impact mitochondrial biogenesis and function through effects on PGC-1α gene expression, which was mentioned above as a target for mHtt.

Mutant Htt has been reported to damage mitochondria by multiple mechanisms: reduced energy production and membrane potential, alteration in calcium homeostasis, increased oxidative stress, and triggering of the caspase-mediated apoptotic pathways (Browne et al., 1997, 1999; Panov et al., 2002).

Htt with as expanded polyglutamine sequesters ubiquitin-proteosome system (UPS) components into Htt inclusions altering their subcellular localization and consequently decaying the UPS efficiency (Davies et al., 1997; DiFiglia et al., 1997). Another method of interfering with the UPS consists on expanded Htt not being degraded efficiently within the proteasome, leading to expanded polyglutamine sequences trapped in the proteasome that may block the entry of other substrates (Imarisio et al., 2008). It has also been

proposed that inefficient degradation of expanded polyglutamine tracts of proteins may lead to the production of aggregation-prone fragments (Holmberg et al., 2004).

Autophagy may also be slightly compromised in some cases. Mutant Htt can interfere with target recognition and impair autophagic clearance (Martinez-Vicente et al., 2010; Qi et al., 2012). Two articles have presented (in animal models) autophagy as a protecting mechanism capable of clearing the cell of mutated forms of Htt (Jia et al., 2007; Ravikumar et al., 2004).

Figure 36 groups the most important pathogenic mechanisms that occur in HD-affected cells.

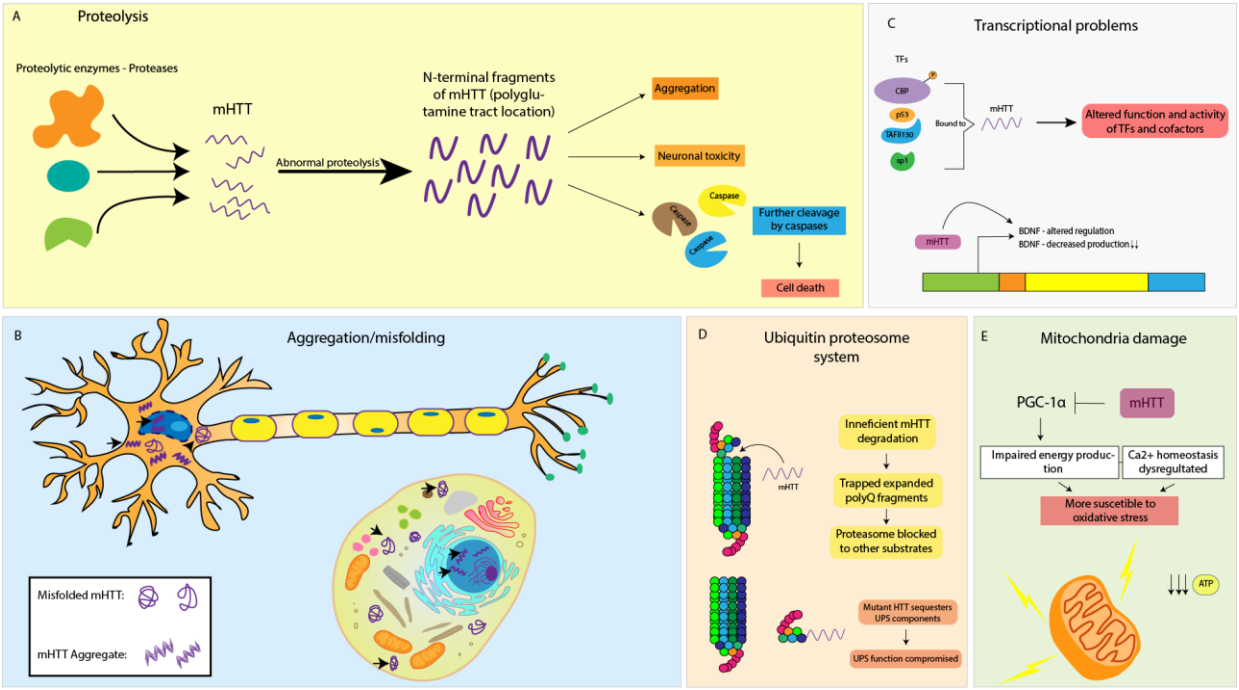

Figure 36: Huntington's disease pathogenesis. A - Proteolysis and further cleavage of mHtt leads to cell toxicity and death. B - Misfolded Htt aggregates populate the cells, although the contribution of this event to the development of the disease remains unclear. C - Mutant Htt binds transcripton factors (TFs) and alters their activity, BDNF levels may decrease as a result of this. D – The ubiquitin-proteosome system (UPS) cannot perform normal degradation on expanded Htt and it might become blocked for other proteins marked for degradation. Also, mHtt can sometimes sequester UPS components, compromising their normal function. E - Mutant Htt affects the expression of PGC-1α, impairing normal mitochondria function leading to increased oxidative stress.

4.4.1.7 Clinical manifestations: neuropathology, signs and symptoms

4.4.1.7.1 Neuropathology

Neuropathologically, HD is primarily characterized by neuronal loss in the cortex and the basal ganglia, especially in the striatum. Other regions usually affected are: globus pallidus (GP), thalamus, hypothalamus, subthalamic nucleus, substantia nigra (SN) and cerebellum (Raymond et al., 2011).

Diffusion tensor imaging has demonstrated pathology of the white matter in pre- and early symptomatic patients (Rosas et al., 2006).

4.4.1.7.2 Signs and symptoms

HD is a progressive disorder that causes motor, cognitive, and psychiatric signs and symptoms. On average, most people begin developing features of HD between ages 35 and 44. Signs and symptoms vary by stage and are listed in table 4.

4.4.2 Spinocerebellar Ataxia type 1

4.4.2.1 Introduction

Spinocerebellar ataxia type 1 (SCA1) is a condition characterized by progressive problems with movement. People with this condition initially experience problems with coordination and balance (ataxia). SCA1 is an autosomal dominant disease, meaning that at least one parent must have the disease for the children to inherit it, and that the odds of any given child inheriting SCA 1, regardless of gender or other phenotypes, is 50% if the affected parent is heterozygous.

Signs and symptoms of the disorder typically begin in early adulthood but can appear anytime from childhood to late adulthood. People with SCA1 typically survive 10 to 20 years after symptoms first appear.

Table 4. Signs and symptoms of Huntington's disease according to disease stages

Early stage	Middle stage	Late stage
Behavioural disturbances	Dystonia	Behavioral disturbances
Clumsiness	Involuntary movements	Rigidity
Moodiness	Trouble with balance and walking	Bradykinesia
Irritability	Chorea with twisting and writhing motions	Severe chorea
Paranoia	Unsteady gait	Serious weight loss
Apathy	Slow reaction time	Inability to speak
Anxiety	General weakness	Inability to walk
Hallucinations	Weight loss	Swallowing problems
Abnormal eye movements	Speech difficulties	Inability to care for oneself
Depression	Stubbornness	
Impaired ability to detect odors		

4.4.2.2 First described

SCA1 research began in the last half of the nineteenth century with genetic applications. Between 1863 and 1877, Friedreich described a recessive form of hereditary ataxia that today bears his name, but in 1893, Pierre Marie noticed a form of ataxia in four families with a clinical picture distinct from that described by Friedreich (Harry T. Orr & Zoghbi, 2001). Ataxias recognized by Marie have evolved into the autosomal dominant ataxias or spinocerebellar ataxias (SCAs).

4.4.2.3 Epidemiology

Worldwide, one to two individuals in 100.000 develop SCA1, with significant geographical and ethnic variations.

Figure 37 presents a world map with the prevalence of SCA1 within the group of autosomal dominant cerebellar ataxias (ADCA), according to GeneReviews (Opal & Ashizawa, 2017).

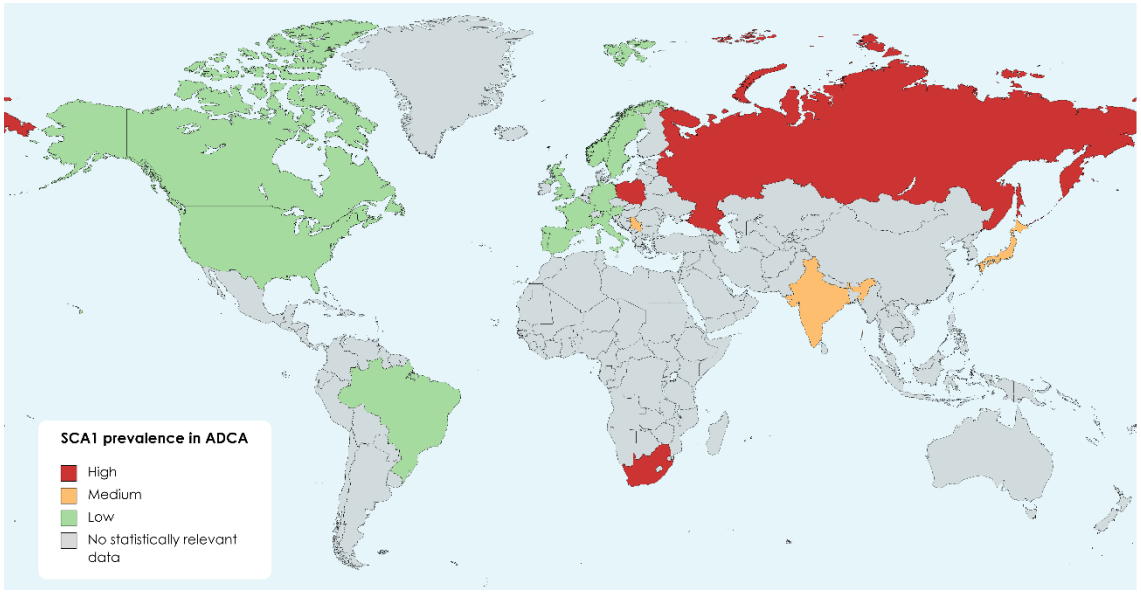

Figure 37: World chart of SCA1 epidemiology based on its prevalence within the group of autosomal dominant cerebellar ataxias

4.4.2.4 Causative gene

The gene associated with SCA1 is termed *ATXN1*, spans 450 kb of genomic DNA and is organized into 9 exons (Banfi et al., 1994). The first 7 fall into the 5-prime untranslated region, whereas the last 2 contain the coding region (21.982 nt) (figure 38). There is an increased ammount of CAG repeats in exon 8 when disease is present. CAG repeats are usually interrupted by CAT trinucleotides (Nethisinghe et al., 2018). CATs are thought to be involved in the stability of the trinucleotide stretch during DNA replication (Nethisinghe et al., 2018).

Normal *ATXN1* alleles contain from 6 to 43 CAG repeats with those longer than 21 repeats being interrupted with 1- 3 CAT trinucleotides. Disease-associated alleles, on the other hand, are CAG tracts ranging from 39 to 44 uninterrupted by CAT trinucleotides. Such interruptions are found in all the longer unaffected alleles. Most affected alleles are pure CAG tracts (Orr, 2012). The presence of repeat interruptions particularly in the longer wild-type alleles lead to the possibility that the CAT interruptions have an important function of maintaining stability of normal alleles (figure 38). ATXN1 gene codes for ataxin-1 and is in the short arm of chromosome 6 (figure 39), according to NCBI.

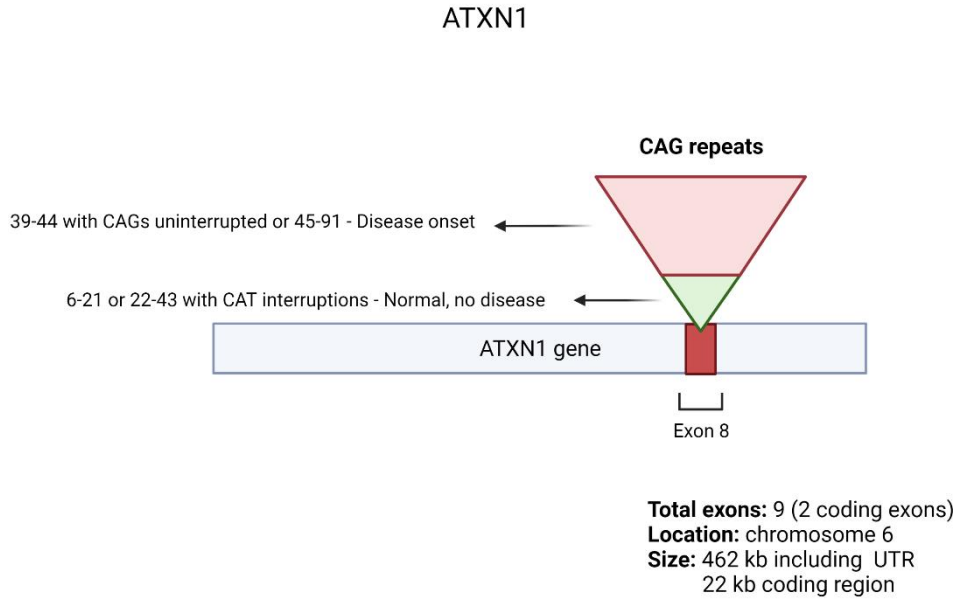

Figure 38: *ATXN1* gene with ranges of the CAG repeat numbers found in SCA1 patients or in healthy individuals, along with information about the total coding exons, chromosome location and size in DNA bases.

4.4.2.5 Codified protein: structure, domains, functions and intracellular localization

4.4.2.5.1 Structure and domains

The *ATXN1* gene codes for ataxin-1. Ataxin-1 has 815 amino acids and a molecular mass of 86.923 Da (~87 kDa). Ataxin-1 is a ubiquitous polyglutamine protein expressed primarily in the nucleus.

SCA1 is caused when ataxin-1 is mutated, this is when ataxin-1 has a bigger than normal number of glutamines, in which case the protein can no longer be stable and leads to cellular problems that eventually result in disease.

An ataxin-1 protein representation is shown in figure 40 along, and table 5 presents information on one of its domains (information from Uniprot.org).

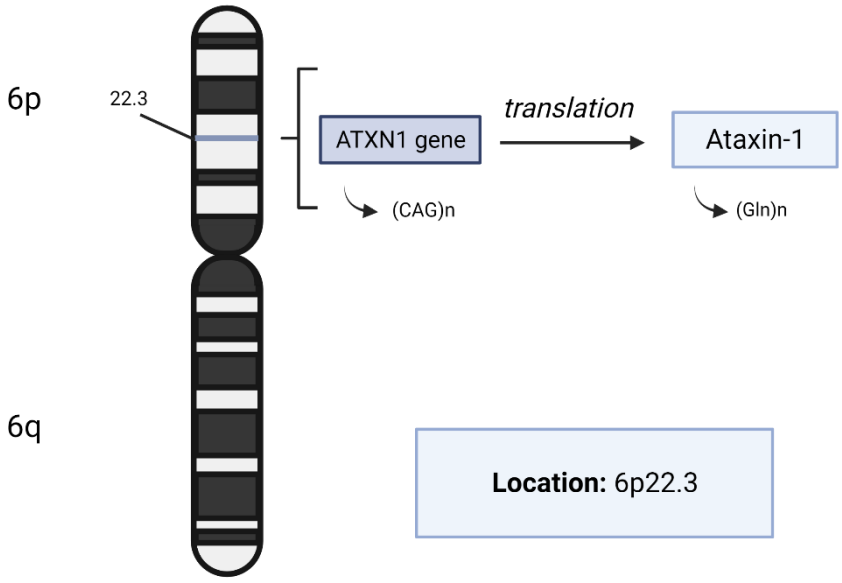

Figure 39: *ATXN1* gene chromosomal band localization.

4.4.2.5.2 Protein functions and biological role

Ataxin-1 has important molecular functions regarding transcription:

- It binds chromatin (DNA) and functions as a transcriptional repressor;

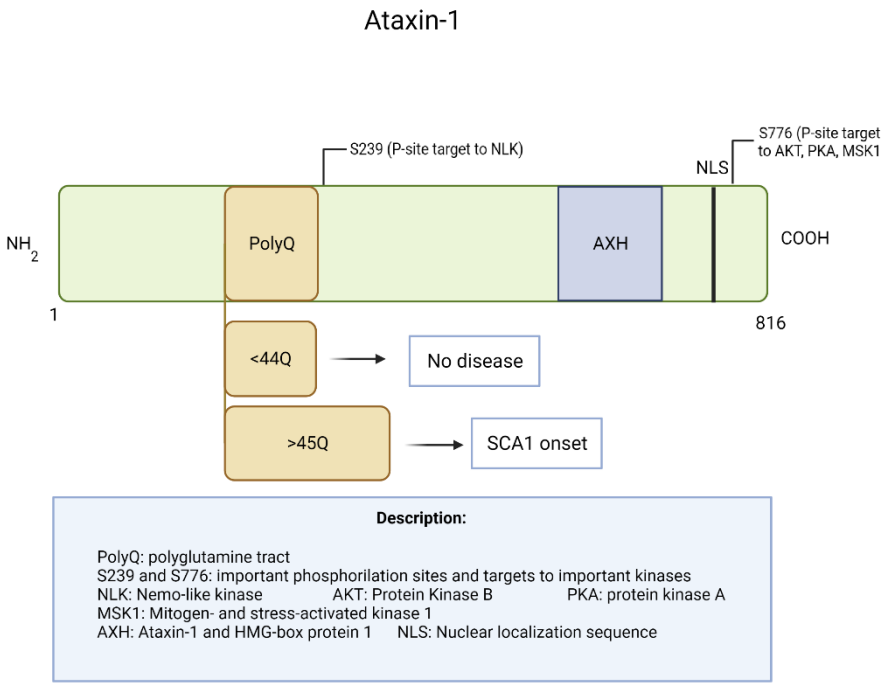

Figure 40: Ataxin-1 domain structure. Caption in figure.

- It can bind RNA (Yue et al., 2001);
- RNA processing function (Irwin et al., 2005).

According to UniProt.org, the *ATXN1* gene product is involved in various biological processes, from early nervous system development to maintaining its stability and normal functioning throughout the years. It has important roles in brain development, learning and memory.

4.4.2.5.3 Protein intracellular localization

Ataxin-1 is widely expressed in the normal human brain and peripheral tissues. It is present in both the nucleus and the cytoplasm. Within neurons, ataxin-1 is predominantly nuclear, with some cytoplasmic ataxin-1 found in Purkinje cells and brainstem nuclei (Opal & Orr, 2015; Servadio et al., 1995).

4.4.2.6 Pathophysiology

Disease-related polyglutamine expansions have a direct relationship between length and severity/age-of-onset of disease, i.e. the longer the glutamine tract the more severe and earlier is the age of onset of disease.

The presence of expanded glutamine repeats in Ataxin-1 impairs ubiquitination and degradation, leading to accumulation of the protein in neurons and subsequent toxicity (Paulson et al., 2017). On the other hand, some other studies reported that accumulation in neurons showed no further disease worsening and disease specimens with less or close to none neuronal inclusions still developed disease, as did cases with accumulation of expanded Ataxin-1 (Klement et al., 1998; Zoghbi & Orr, 2009).

Table 5. **Ataxin-1 important domain**

Description	**Position**	**Length**
AXH	562-689	127

Biochemical and genetic studies provide evidence that the polyglutamine expansion enhances interactions that are normally regulated by phosphorylation at S776 and induces a subsequent alteration in its interaction with other cellular proteins (Zoghbi & Orr, 2009).

S776 phosphorylation confers great stability to Ataxin-1 and enhances protein-protein interactions. Interaction of pS776-Ataxin-1 with 14-3-3 stabilizes the first. The chaperone 14-3-3 bound to pS776-Ataxin-1 protects the latter from degradation, which may lead to toxic levels of the protein (Nóbrega & Almeida, 2018).

SUMOylation is a post-translational modification involved in various cellular processes, such as nuclear-cytosolic transport, transcriptional regulation, apoptosis, protein stability, response to stress, and progression through the cell cycle. SUMOylation of Ataxin-1 is dependent on nuclear localization and phosphorylation at Ser-776. It is reduced in the presence of an expanded polyglutamine tract (Riley et al., 2005; Zoghbi & Orr, 2009).

The Ataxin-1 AXH domain allows its homodimerization and interaction with regulators of transcription such as the transcriptional repressor Capicua (Cic). Ataxin-1 and Cic together exist in large protein complexes where mutant Ataxin-1 is thought to affect Cic repressor activity. In addition to regulating Cic function, stability of Cic depends on Ataxin-1 binding (Nóbrega & Pereira de Almeida, 2018). Expanded Ataxin-1 displays an enhanced interaction with RNA-binding motif protein 17 (RBM17), since RBM17 interacted preferably with polyglutamine-expanded Ataxin-1 and its phosphorylated forms.

Expanded Ataxin-1 toxicity results from a gain of function mechanism, where mutant Ataxin-1 affects the normal function of an Ataxin-1 binding partner (Cic) and exacerbates its normal gene-repressor function *in vivo* (Lam et al., 2006; Rousseaux et al., 2018). However, mutant Ataxin-1 also produces a Cic loss of function, something that reveals the complexity of mutant Ataxin-1 effects (Rousseaux et al., 2018).

Figure 41 groups the most important pathogenic mechanisms that occur in SCA1-affected cells.

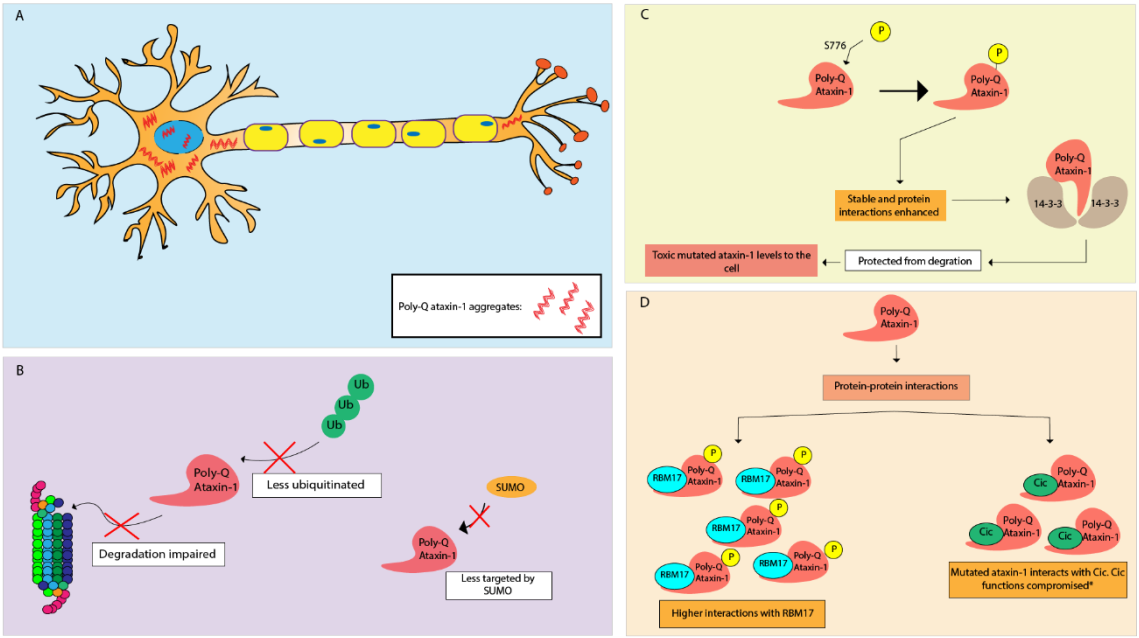

Figure 41: SCA1 pathogenesis. A - Aggregation of mutated Ataxin-1 inside the cell and nucleus. B - Expanded Ataxin-1 is less ubiquitinated than wild-type protein, leading to less degradation of the mutated form. Ataxin-1 with expanded glutamines is also less targeted by SUMO and so it is less SUMOylated, affecting cellular stability. C - Interaction of mutated Ataxin-1 with chaperones is enhanced, resulting in the protection of the first from degradation therefore increasing toxic levels of the mutated protein inside the cell. D - Expanded ataxin-1 shifts the balance of interactions. Mutated Ataxin-1 interacts more with RBM17 and affects Cic function leading to its exacerbated function or loss of function, in some cases.

4.4.2.7 Clinical manifestations: neuropathology, signs and symptoms

4.4.2.7.1 Neuropathology

SCA1 primarily affects the brainstem, spinocerebellar tracts and cerebellar Purkinje cells.

SCA1 is associated with many pathoanatomical findings in various tissues (Rüb et al., 2012):

- Neuronal loss in the neocortex
- Neuronal loss in the basal forebrain nuclei, basal ganglia, amygdala, hypothalamus and thalamus
- Neuronal loss in the cerebellum
- Neuronal loss in the midbrain
- Neuronal loss in the pons and medulla oblongata

- Neuronal intranuclear inclusion
- Astrogliosis

4.4.2.7.2 Signs and symptoms

Table 6 presents a list of neurological signs and symptoms and other manifestation often seen in SCA1 patients.

Table 6. Early and later symptoms of SCA1.

Early symptoms	Later symptoms
Speech and swallowing difficulties	Dystonia
Ataxia of gait	Atrophy
Ataxia of stance	Sensory neuropathy
Spasticity	Fasciculations
Nystagmus	Rigidity
Ophthalmoplegia	Tremors
Cognitive impairment	

CHAPTER 5 - DISCUSSION

5. DISCUSSION

5.1 PolyQ Database accomplishments

This project accomplished all the predefined objectives with added functionalities. The PolyQ Database created not only presents all the information discussed in previous chapters, it also offers simple functionalities to navigate the platform.

By using a database to store information of all polyQ diseases, any change to text or media can be easily updated in real-time with no need for further coding. This fast method of updating the database stored values changes (almost instantly) the information present in each of the Disease pages.

Another important factor to mention is platform responsibility. The PolyQ Database is compatible with all resolutions used in computers, tablets and phone mobile devices. This allows for proper visualization on every device.

Finally, the PolyQ Database platform was made to ensure new Disease pages can be included on the possibility that new polyQ diseases are described.

5.2 PolyQ database comparison with other databases

The PolyQ Database, as many other databases such as OMIM and Orphanet, follows a structure for each entry, in this case for each polyQ disease. Most of the databases offer only textual information regarding any point they focus on. The PolyQ Database, although it does also offer significant amounts of textual information, accompanies the text with supplementary images, which constitute an appealing visual support. Compared to other polyQ databases, our database does not focus only on a few specific topics. The goal was always to develop an online resource that covers all the fundamental disease information. Another difference is perhaps the simplicity. Accessing the polyQ disease of interest is straightforward. Compared to some other databases where the search process is rather complicated, with detailed explanations for the search mechanism, the PolyQ Database only requires the user to know the name of the disease, and if that is unknown,

there are help messages and the current available diseases have quick shortcuts that do not requiring any search.

The PolyQ Database is user friendly and is responsive to any used device and resolution. This versatility is sometimes not seen in older databases.

5.3 PolyQ Database problems and challenges

The PolyQ Database was a project implemented with tools that were not something usually used by a student in a Biomedical Sciences degree. Programming and understanding of database structure are not necessary skill to be found in a Biology course. Although there was some minor previous understanding of programming, many of the tools used were entirely new and learned from scratch. The website back-end had some major changes in late stages of development because of encountered issues, which postponed some features. Back-end was by far the hardest part to implement from a technical point of view, but thankfully there was help from a collaborator with professional background in the area.

Because of time limitations, we were unable to redo some features and improve future proofing. Perhaps in time those changes may occur and further improve the website management. Although there were issues, the final product accomplished most of what we had in mind.

5.4 Disease topics. What was not added.

Initially, the information topics presented for each disease were not exactly structured as they are presented in the final product. Therapeutic approaches for each disease were initially planned to be included as well, but it was decided that this would not be implemented. One very important matter that could be presented was the subject of biomarkers that exist for each disease, but there is no available information for every polyQ disease regarding this aspect, and therefore it would not make sense to have this topic added to a set of diseases (4 or 6) and not all.

The Disease pages benefit from having a clear line of thought from top to bottom. These pages present complex disease mechanisms and cell functions, but overall the information provided establishes the appropriate background and links to other sources that contain further information. Overall, the information provides a solid base for anyone interested in the field of polyQ diseases and/or that is searching for data for further studies.

5.5 Conflicting data across various sources

During the data collection step of this project, there were various instances where the information obtained was not consistent. This part of the project ended up being a bibliographic review for all nine polyQ diseases. On more technical matters such as protein structure and pathophysiological mechanisms there were several studies with different conclusions, that never neglected others. Although research efforts have provided extensive information on polyQ diseases, various teams are still working on deepening our knowledge about these complex human conditions, in order to contribute to the future development of therapies. The PolyQ Database offers current solid evidence-based data on these disorders, curated from research work published by various authors around the world.

CHAPTER 6 – CONCLUSIONS AND FUTURE PERSPECTIVES

6. CONCLUSIONS AND FUTURE PERSPECTIVES

PolyQ diseases have been initially described decades ago, and although there are various online sources and studies that contain information about them, there are very few resources that exclusively focus on this group of neurodegenerative diseases. The PolyQ Database is the first polyQ disease online resource that focuses exclusively on the fundamentals and mechanisms of every polyQ disorder. This project provides a new web-based tool which will be useful for researchers, students or any other visitor.

As mentioned before, regarding platform dynamicity, all the information is easily updated. New discoveries that alter the current paradigms can be rapidly included, and other aspects are easily changed thanks to the back-end developed for the database.

With time, there will be major updates to the website “behind the scenes”, which will require new strategies. The back-end navigation will be further improved for better understanding of the logic behind it so other individuals responsible for maintenance do not struggle. Besides doing improvements to the back-end, additional features that will enable further exploration of gene and protein information will eventually be explored and implemented.

REFERENCES

Arrasate, M., & Finkbeiner, S. (2012). Protein aggregates in Huntington's disease. *Experimental Neurology*, *238*(1), 1–11. https://doi.org/10.1016/J.EXPNEUROL.2011.12.013

Arrasate, M., Mitra, S., Schweitzer, E. S., Segal, M. R., & Finkbeiner, S. (2004). Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. *Nature*, *431*(7010), 805–810. https://doi.org/10.1038/NATURE02998

Bacolla, A., Jaworski, A., Larson, J. E., Jakupciak, J. P., Chuzhanova, N., Abeysinghe, S. S., O'Connell, C. D., Cooper, D. N., & Wells, R. D. (2004). Breakpoints of gross deletions coincide with non-B DNA conformations. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(39), 14162. https://doi.org/10.1073/PNAS.0405974101

Bacolla, A., & Wells, R. D. (2004). Non-B DNA Conformations, Genomic Rearrangements, and Human Disease *. *Journal of Biological Chemistry*, *279*(46), 47411–47414. https://doi.org/10.1074/JBC.R400028200

Bai, J. J., Safadi, S. S., Mercier, P., Barber, K. R., & Shaw, G. S. (2013). Ataxin-3 is a multivalent ligand for the parkin Ubl domain. *Biochemistry*, *52*(42), 7369–7376. https://doi.org/10.1021/BI400780V

Banfi, S., Servadio, A., Chung, M. yi, Kwiatkowski, T. J., McCall, A. E., Duvick, L. A., Shen, Y., Roth, E. J., Orr, H. T., & Zoghbi, H. Y. (1994). Identification and characterization of the gene causing type 1 spinocerebellar ataxia. *Nature Genetics 1994 7:4*, *7*(4), 513–520. https://doi.org/10.1038/ng0894-513

Bernardi, K. M., Williams, J. M., Inoue, T., Schultz, A., & Tsai, B. (2013). A deubiquitinase negatively regulates retro-translocation of nonubiquitinated substrates. *Molecular Biology of the Cell*, *24*(22), 3545–3556. https://doi.org/10.1091/MBC.E13-06-0332

Bettencourt, C., & Lima, M. (2011). Machado-Joseph disease: From first descriptions to new perspectives. *Orphanet Journal of Rare Diseases*, *6*(1), 1–12. https://doi.org/10.1186/1750-1172-6-35/TABLES/1

Bhide, P. G., Day, M., Sapp, E., Schwarz, C., Sheth, A., Kim, J., Young, A. B., Penney, J., Golden, J., Aronin, N., & DiFiglia, M. (1996). Expression of normal and mutant huntingtin in the developing brain. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *16*(17), 5523–5535. https://doi.org/10.1523/JNEUROSCI.16-17-05523.1996

Browne, S. E., Bowling, A. C., MacGarvey, U., Baik, M. J., Berger, S. C., Muqit, M. M. K., Bird, E. D., & Beal, M. F. (1997). Oxidative damage and metabolic dysfunction in Huntington's disease: selective vulnerability of the basal ganglia. *Annals of Neurology*, *41*(5), 646–653. https://doi.org/10.1002/ANA.410410514

Browne, S. E., Ferrante, R. J., & Beal, M. F. (1999). Oxidative stress in Huntington's disease. *Brain Pathology (Zurich, Switzerland)*, *9*(1), 147–163. https://doi.org/10.1111/J.1750-3639.1999.TB00216.X

Budworth, H., & McMurray, C. T. (2013a). A Brief History of Triplet Repeat Diseases. *Methods in Molecular Biology (Clifton, N.J.)*, *1010*, 3. https://doi.org/10.1007/978-1-62703-411-1_1

Budworth, H., & McMurray, C. T. (2013b). A Brief History of Triplet Repeat Diseases. *Methods in Molecular Biology (Clifton, N.J.), 1010*, 3. https://doi.org/10.1007/978-1-62703-411-1_1

Campos-Romo, A., Graue-Hernandez, E. O., Pedro-Aguilar, L., Hernandez-Camarena, J. C., Rivera-De La Parra, D., Galvez, V., Diaz, R., Jimenez-Corona, A., & Fernandez-Ruiz, J. (2017). Ophthalmic features of spinocerebellar ataxia type 7. *Eye 2018 32:1*, *32*(1), 120–127. https://doi.org/10.1038/eye.2017.135

Caron, N. S., Wright, G. E., & Hayden, M. R. (2020). Huntington Disease. *GeneReviews®*. https://www.ncbi.nlm.nih.gov/books/NBK1305/

Casey, H. L., & Gomez, C. M. (2019). *Spinocerebellar Ataxia Type 6*. *54*, 473–479. https://www.ncbi.nlm.nih.gov/books/NBK1140/

Casorelli, I., Bossa, C., & Bignami, M. (2012). DNA Damage and Repair in Human Cancer: Molecular Mechanisms and Contribution to Therapy-Related Leukemias. *International Journal of Environmental Research and Public Health*, *9*(8), 2636. https://doi.org/10.3390/IJERPH9082636

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B. M., Strein, C., Davey, N. E., Humphreys, D. T., Preiss, T., Steinmetz, L. M., Krijgsveld, J., & Hentze, M. W. (2012). Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell*, *149*(6), 1393–1406. https://doi.org/10.1016/J.CELL.2012.04.031

Chen, S., Ferrone, F. A., & Wetzel, R. (2002). Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(18), 11884. https://doi.org/10.1073/PNAS.182276099

Chen, X., Tang, T. S., Tu, H., Nelson, O., Pook, M., Hammer, R., Nukina, N., & Bezprozvanny, I. (2008). Deranged calcium signaling and neurodegeneration in spinocerebellar ataxia type 3. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *28*(48), 12713–12724. https://doi.org/10.1523/JNEUROSCI.3909-08.2008

Chou, A. H., Yeh, T. H., Kuo, Y. L., Kao, Y. C., Jou, M. J., Hsu, C. Y., Tsai, S. R., Kakizuka, A., & Wang, H. L. (2006). Polyglutamine-expanded ataxin-3 activates mitochondrial apoptotic pathway by upregulating Bax and downregulating Bcl-xL. *Neurobiology of Disease*, *21*(2), 333–345. https://doi.org/10.1016/J.NBD.2005.07.011

Cornelius, N., Wardman, J. H., Hargreaves, I. P., Neergheen, V., Bie, A. S., Tümer, Z., Nielsen, J. E., & Nielsen, T. T. (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*, 103–114. https://doi.org/10.1016/J.MITO.2017.03.001

Cortes, C. J., & La Spada, A. R. (2015). Autophagy in Polyglutamine Disease: Imposing Order on Disorder or Contributing to the Chaos? *Molecular and Cellular Neurosciences*, *66*(0 0), 53. https://doi.org/10.1016/J.MCN.2015.03.010

CSS: Cascading Style Sheets | MDN. (n.d.). Retrieved August 19, 2022, from https://developer.mozilla.org/en-US/docs/Web/CSS

Cui, L., Jeong, H., Borovecki, F., Parkhurst, C. N., Tanese, N., & Krainc, D. (2006). Transcriptional repression of PGC-1alpha by mutant huntingtin leads to mitochondrial dysfunction and neurodegeneration. *Cell*, *127*(1), 59–69. https://doi.org/10.1016/J.CELL.2006.09.015

Database Fundamentals #8: All About Data Types - DZone Database. (n.d.). Retrieved July 26, 2022, from https://dzone.com/articles/database-fundamentals-8-all-about-data-types

Database Management System (DBMS). (n.d.). Retrieved July 26, 2022, from https://www.techtarget.com/searchdatamanagement/definition/database-management-system

Davies, S. W., Turmaine, M., Cozens, B. A., DiFiglia, M., Sharp, A. H., Ross, C. A., Scherzinger, E., Wanker, E. E., Mangiarini, L., & Bates, G. P. (1997). Formation of neuronal intranuclear inclusions underlies the neurological dysfunction in mice transgenic for the HD mutation. *Cell*, *90*(3), 537–548. https://doi.org/10.1016/S0092-8674(00)80513-9

DBMS Architecture - Detailed Explanation - InterviewBit. (n.d.). Retrieved August 11, 2022, from https://www.interviewbit.com/blog/dbms-architecture/

DBMS Architecture - javatpoint. (n.d.). Retrieved August 11, 2022, from https://www.javatpoint.com/dbms-architecture

de Koning, A. P. J., Gu, W., Castoe, T. A., Batzer, M. A., & Pollock, D. D. (2011). Repetitive Elements May Comprise Over Two-Thirds of the Human Genome. *PLOS Genetics*, *7*(12), e1002384. https://doi.org/10.1371/JOURNAL.PGEN.1002384

DiFiglia, M., Sapp, E., Chase, K. O., Davies, S. W., Bates, G. P., Vonsattel, J. P., & Aronin, N. (1997). Aggregation of huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain. *Science (New York, N.Y.)*, *277*(5334), 1990–1993. https://doi.org/10.1126/SCIENCE.277.5334.1990

DiFiglia, M., Sapp, E., Chase, K., Schwarz, C., Meloni, A., Young, C., Martin, E., Vonsattel, J. P., Carraway, R., Reeves, S. A., Boyce, F. M., & Aronin, N. (1995). Huntingtin is a cytoplasmic protein associated with vesicles in human and rat brain neurons. *Neuron*, *14*(5), 1075–1081. https://doi.org/10.1016/0896-6273(95)90346-1

Dion, V., Lin, Y., Hubert, L., Waterland, R. A., & Wilson, J. H. (2008). Dnmt1 Deficiency Promotes CAG Repeat Expansion in the Mouse Germline. *Human Molecular Genetics*, *17*(9), 1306. https://doi.org/10.1093/HMG/DDN019

Dunah, A. W., Jeong, H., Griffin, A., Kim, Y. M., Standaert, D. G., Hersch, S. M., Mouradian, M. M., Young, A. B., Tanese, N., & Krainc, D. (2002). Sp1 and TAFII130 transcriptional activity disrupted in early Huntington's disease. *Science (New York, N.Y.)*, *296*(5576), 2238–2243. https://doi.org/10.1126/SCIENCE.1072613

Evers, M. M., Toonen, L. J. A., & Van Roon-Mom, W. M. C. (2014). Ataxin-3 Protein and RNA Toxicity in Spinocerebellar Ataxia Type 3: Current Insights and Emerging Therapeutic Strategies. *Molecular Neurobiology*, *49*(3), 1513. https://doi.org/10.1007/S12035-013-8596-2

Filla, A., Mariotti, C., Caruso, G., Coppola, G., Cocozza, S., Castaldo, I., Calabrese, O., Salvatore, E., De Michele, G., Riggio, M. C., Pareyson, D., Gellera, C., & Di Donato, S. (2000). Relative frequencies of CAG expansions in spinocerebellar ataxia and dentatorubropallidoluysian atrophy in 116 Italian families. *European Neurology*, *44*(1), 31–36. https://doi.org/10.1159/000008189

Friedman, M. J., Shah, A. G., Fang, Z. H., Ward, E. G., Warren, S. T., Li, S., & Li, X. J. (2007). Polyglutamine domain modulates the TBP-TFIIB interaction: implications for its normal function and neurodegeneration. *Nature Neuroscience*, *10*(12), 1519–1528. https://doi.org/10.1038/NN2011

Friedman, M. J., Wang, C. E., Li, X. J., & Li, S. (2008). Polyglutamine expansion reduces the association of TATA-binding protein with DNA and induces DNA binding-independent neurotoxicity. *The Journal of Biological Chemistry*, *283*(13), 8283–8290. https://doi.org/10.1074/JBC.M709674200

Friedrich, J. K., Panov, K. I., Cabart, P., Russell, J., & Zomerdijk, J. C. B. M. (2005). TBP-TAF complex SL1 directs RNA polymerase I pre-initiation complex formation and stabilizes upstream binding factor at the rDNA promoter. *The Journal of Biological Chemistry*, *280*(33), 29551–29558. https://doi.org/10.1074/JBC.M501595200

Froment J, Bonnet P, C. A. (1937). Heredo-degenerations retinienne et spino-cerebelleuse: variantes ophtalmoscopiques et neurologiques presentees par trois generations successives. *Med Lyon*, 153–163.

Front End vs Back End Development: What's the Difference? - Kenzie Academy. (n.d.). Retrieved August 18, 2022, from https://kenzie.snhu.edu/blog/front-end-vs-back-end-whats-the-difference/

Fujioka, S., Sundal, C., & Wszolek, Z. K. (2013). Autosomal dominant cerebellar ataxia type III: a review of the phenotypic and genotypic characteristics. *Orphanet Journal of Rare Diseases*, *8*, 14. https://doi.org/10.1186/1750-1172-8-14

Functions of Database Management System (DBMS) - QS Study. (n.d.). Retrieved July 26, 2022, from https://qsstudy.com/functions-of-database-management-system-dbms/

Gatchel, J. R., & Zoghbi, H. Y. (2005). Diseases of unstable repeat expansion: mechanisms and common principles. *Nature Reviews. Genetics*, *6*(10), 743–755. https://doi.org/10.1038/NRG1691

Gauthier, L. R., Né Dicte, B., Charrin, C., Borrell-Pagè, M., Dompierre, J. P., Lè Ne Rangone, H., Cordeliè Res, F. P., De Mey, J., Macdonald, M. E., Leßmann, V., & Humbert, S. (2004). Huntingtin Controls Neurotrophic Support and Survival of Neurons by Enhancing BDNF Vesicular Transport along Microtubules. *Cell*, *118*, 127–138. http://www.cell.com/cgi/content/

Gervais, F. G., Singaraja, R., Xanthoudakis, S., Gutekunst, C. A., Leavitt, B. R., Metzler, M., Hackam, A. S., Tam, J., Vaillancourt, J. P., Houtzager, V., Rasper, D. M., Roy, S., Hayden, M. R., & Nicholson, D. W. (2002). Recruitment and activation of caspase-8 by the Huntingtin-interacting protein Hip-1 and a novel partner Hippi. *Nature Cell Biology 2002 4:2*, *4*(2), 95–105. https://doi.org/10.1038/ncb735

Getting Started with Atom | Codecademy. (n.d.). Retrieved August 19, 2022, from

https://www.codecademy.com/article/f1-text-editors

Goldberg, Y. P., Nicholson, D. W., Rasper, D. M., Kalchman, M. A., Koide, H. B., Graham, R. K., Bromm, M., Kazemi-Esfarjani, P., Thornberry, N. A., Vaillancourt, J. P., & Hayden, M. R. (1996). Cleavage of huntingtin by apopain, a proapoptotic cysteine protease, is modulated by the polyglutamine tract. *Nature Genetics*, *13*(4), 442–449. https://doi.org/10.1038/NG0896-442

Goldin, A. L., Barchi, R. L., Caldwell, J. H., Hofmann, F., Howe, J. R., Hunter, J. C., Kallen, R. G., Mandel, G., Meisler, M. H., Netter, Y. B., Noda, M., Tamkun, M. M., Waxman, S. G., Wood, J. N., & Catterall, W. A. (2000). Nomenclature of voltage-gated sodium channels. *Neuron*, *28*(2), 365–368. https://doi.org/10.1016/S0896-6273(00)00116-1

Gouge, J., Satia, K., Guthertz, N., Widya, M., Thompson, A. J., Cousin, P., Dergai, O., Hernandez, N., & Vannini, A. (2015). Redox Signaling by the RNA Polymerase III TFIIB-Related Factor Brf2. *Cell*, *163*(6), 1375–1387. https://doi.org/10.1016/J.CELL.2015.11.005

Graham, R. K., Deng, Y., Slow, E. J., Haigh, B., Bissada, N., Lu, G., Pearson, J., Shehadeh, J., Bertram, L., Murphy, Z., Warby, S. C., Doty, C. N., Roy, S., Wellington, C. L., Leavitt, B. R., Raymond, L. A., Nicholson, D. W., & Hayden, M. R. (2006). Cleavage at the caspase-6 site is required for neuronal dysfunction and degeneration due to mutant huntingtin. *Cell*, *125*(6), 1179–1191. https://doi.org/10.1016/J.CELL.2006.04.026

Hackam, A. S., Singaraja, R., Wellington, C. L., Metzler, M., McCutcheon, K., Zhang, T., Kalchman, M., & Hayden, M. R. (1998). The Influence of Huntingtin Protein Size on Nuclear Localization and Cellular Toxicity. *The Journal of Cell Biology*, *141*(5), 1097. https://doi.org/10.1083/JCB.141.5.1097

Halliday, G. M., McRitchie, D. A., Macdonald, V., Double, K. L., Trent, R. J., & McCusker, E. (1998). Regional specificity of brain atrophy in Huntington's disease. *Experimental Neurology*, *154*(2), 663–672. https://doi.org/10.1006/EXNR.1998.6919

Hegazy, Y. A., Fernando, C. M., & Tran, E. J. (2020). The balancing act of R-loop biology: The good, the bad, and the ugly. *The Journal of Biological Chemistry*, *295*(4), 905. https://doi.org/10.1074/JBC.REV119.011353

Hegde, M. V., & Saraph, A. A. (2011). Unstable genes unstable mind: beyond the central dogma of molecular biology. *Medical Hypotheses*, *77*(2), 165–170. https://doi.org/10.1016/J.MEHY.2011.03.051

Helmlinger, D., Hardy, S., Abou-Sleymane, G., Eberlin, A., Bowman, A. B., Gansmüller, A., Picaud, S., Zoghbi, H. Y., Trottier, Y., Tora, L., & Devys, D. (2006). Glutamine-expanded ataxin-7 alters TFTC/STAGA recruitment and chromatin structure leading to photoreceptor dysfunction. *PLoS Biology*, *4*(3), 0432–0445. https://doi.org/10.1371/JOURNAL.PBIO.0040067

Helmlinger, D., Hardy, S., Sasorith, S., Klein, F., Robert, F., Weber, C., Miguet, L., Potier, N., Van-Dorsselaer, A., Wurtz, J. M., Mandel, J. L., Tora, L., & Devys, D. (2004). Ataxin-7 is a subunit of GCN5 histone acetyltransferase-containing complexes. *Human Molecular Genetics*, *13*(12), 1257–1265.

https://doi.org/10.1093/HMG/DDH139

Her, L.-S., & Goldstein, L. S. B. (2008). *Enhanced Sensitivity of Striatal Neurons to Axonal Transport Defects Induced by Mutant Huntingtin*. https://doi.org/10.1523/JNEUROSCI.4144-08.2008

Hoffmann, A., Sinn, E., Yamamoto, T., Wang, J., Roy, A., Horikoshi, M., & Roeder, R. G. (1990). Highly conserved core domain and unique N terminus with presumptive regulatory motifs in a human TATA factor (TFIID). *Nature*, *346*(6282), 387–390. https://doi.org/10.1038/346387A0

Hoffner, G., Kahlem, P., & Djian, P. (2002). Perinuclear localization of huntingtin as a consequence of its binding to microtubules through an interaction with beta-tubulin: relevance to Huntington's disease. *Journal of Cell Science*, *115*(Pt 5), 941–948. https://doi.org/10.1242/JCS.115.5.941

Holmberg, C. I., Staniszewski, K. E., Mensah, K. N., Matouschek, A., & Morimoto, R. I. (2004). Inefficient degradation of truncated polyglutamine proteins by the proteasome. *The EMBO Journal*, *23*(21), 4307–4318. https://doi.org/10.1038/SJ.EMBOJ.7600426

Hsiao, P. W., & Chang, C. (1999). Isolation and characterization of ARA160 as the first androgen receptor N-terminal-associated coactivator in human prostate cells. *The Journal of Biological Chemistry*, *274*(32), 22373–22379. https://doi.org/10.1074/JBC.274.32.22373

Hsu, T. C., Wang, C. K., Yang, C. Y., Lee, L. C., Hsieh-Li, H. M., Ro, L. S., Chen, C. M., Lee-Chen, G. J., & Su, M. T. (2014). Deactivation of TBP contributes to SCA17 pathogenesis. *Human Molecular Genetics*, *23*(25), 6878–6893. https://doi.org/10.1093/HMG/DDU410

Huang, S., Ling, J. J., Yang, S., Li, X. J., & Li, S. (2011). Neuronal expression of TATA box-binding protein containing expanded polyglutamine in knock-in mice reduces chaperone protein response by impairing the function of nuclear factor-Y transcription factor. *Brain : A Journal of Neurology*, *134*(Pt 7), 1943–1958. https://doi.org/10.1093/BRAIN/AWR146

Huang, S., Zhu, S., Li, X. J., & Li, S. (2019). The Expanding Clinical Universe of Polyglutamine Disease. *The Neuroscientist : A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, *25*(5), 512. https://doi.org/10.1177/1073858418822993

Hughes, A., & Jones, L. (2011). Huntingtin localisation studies - a technical review. *PLoS Currents*, *3*. https://doi.org/10.1371/CURRENTS.RRN1211

HyperText Markup Language – HTML Definition. (n.d.). Retrieved August 19, 2022, from https://www.investopedia.com/terms/h/html.asp

Imarisio, S., Carmichael, J., Korolchuk, V., Chen, C. W., Saiki, S., Rose, C., Krishna, G., Davies, J. E., Ttofi, E., Underwood, B. R., & Rubinsztein, D. C. (2008). Huntington's disease: from pathology and genetics to potential therapies. *The Biochemical Journal*, *412*(2), 191–209. https://doi.org/10.1042/BJ20071619

Irwin, S., Vandelft, M., Pinchev, D., Howell, J. L., Graczyk, J., Orr, H. T., & Truant, R. (2005). RNA association and nucleocytoplasmic shuttling by ataxin-1. *Journal of*

Cell Science, *118*(Pt 1), 233–242. https://doi.org/10.1242/JCS.01611

Jia, K., Hart, A. C., & Levine, B. (2007). Autophagy genes protect against disease caused by polyglutamine expansion proteins in Caenorhabditis elegans. *Autophagy*, *3*(1), 21–25. https://doi.org/10.4161/AUTO.3528

Jimenez-Sanchez, M., Licitra, F., Underwood, B. R., & Rubinsztein, D. C. (2017). Huntington's Disease: Mechanisms of Pathogenesis and Therapeutic Strategies. *Cold Spring Harbor Perspectives in Medicine*, *7*(7), 1–22. https://doi.org/10.1101/CSHPERSPECT.A024240

Kay, C., Hayden, M. R., & Leavitt, B. R. (2017). Epidemiology of Huntington disease. *Handbook of Clinical Neurology*, *144*, 31–46. https://doi.org/10.1016/B978-0-12-801893-4.00003-1

Kegel, K. B., Meloni, A. R., Yi, Y., Kim, Y. J., Doyle, E., Cuiffo, B. G., Sapp, E., Wang, Y., Qin, Z. H., Don Chen, J., Nevins, J. R., Aronin, N., & Difiglia, M. (2002). Huntingtin is present in the nucleus, interacts with the transcriptional corepressor C-terminal binding protein, and represses transcription. *The Journal of Biological Chemistry*, *277*(9), 7466–7476. https://doi.org/10.1074/JBC.M103946200

Kennedy, W. R., Alter, M., & Sung, J. H. (1968). Progressive proximal spinal and bulbar muscular atrophy of late onset. *Neurology*, *18*(7), 671–671. https://doi.org/10.1212/WNL.18.7.671

Kerrest, A., Anand, R. P., Sundararajan, R., Bermejo, R., Liberi, G., Dujon, B., Freudenreich, C. H., & Richard, G. F. (2009). SRS2 and SGS1 prevent chromosomal breaks and stabilize triplet repeats by restraining recombination. *Nature Structural & Molecular Biology*, *16*(2), 159. https://doi.org/10.1038/NSMB.1544

Kinney, N., Kang, L., Eckstrand, L., Pulenthiran, A., Samuel, P., Anandakrishnan, R., Varghese, R. T., Michalak, P., & Garner, H. R. (2019). Abundance of ethnically biased microsatellites in human gene regions. *PLOS ONE*, *14*(12), e0225216. https://doi.org/10.1371/JOURNAL.PONE.0225216

Klement, I. A., Skinner, P. J., Kaytor, M. D., Yi, H., Hersch, S. M., Clark, H. B., Zoghbi, H. Y., & Orr, H. T. (1998). Ataxin-1 nuclear localization and aggregation: role in polyglutamine-induced disease in SCA1 transgenic mice. *Cell*, *95*(1), 41–53. https://doi.org/10.1016/S0092-8674(00)81781-X

Klockgether, T. (2007). Ataxias. *Textbook of Clinical Neurology: Third Edition*, 765–780. https://doi.org/10.1016/B978-141603618-0.10035-9

Kordasiewicz, H. B., Thompson, R. M., Clark, H. B., & Gomez, C. M. (2006). C-termini of P/Q-type Ca2+ channel alpha1A subunits translocate to nuclei and promote polyglutamine-mediated toxicity. *Human Molecular Genetics*, *15*(10), 1587–1599. https://doi.org/10.1093/HMG/DDL080

Koshy, B. T., & Zoghbi, H. Y. (1997). The CAG/Polyglutamine Tract Diseases: Gene Products and Molecular Pathogenesis. *Brain Pathology*, *7*(3), 927. https://doi.org/10.1111/J.1750-3639.1997.TB00894.X

Kovtun, I. V., Thornhill, A. R., & McMurray, C. T. (2004). Somatic deletion events occur during early embryonic development and modify the extent of CAG expansion in subsequent generations. *Human Molecular Genetics*, *13*(24), 3057–3068.

https://doi.org/10.1093/HMG/DDH325

Kovtun, Irina V., Liu, Y., Bjoras, M., Klungland, A., Wilson, S. H., & McMurray, C. T. (2007). OGG1 initiates age-dependent CAG trinucleotide expansion in somatic cells. *Nature*, *447*(7143), 447. https://doi.org/10.1038/NATURE05778

Kubodera, T., Yokota, T., Ohwada, K., Ishikawa, K., Miura, H., Matsuoka, T., & Mizusawa, H. (2003). Proteolytic cleavage and cellular toxicity of the human alpha1A calcium channel in spinocerebellar ataxia type 6. *Neuroscience Letters*, *341*(1), 74–78. https://doi.org/10.1016/S0304-3940(03)00156-3

La Spada, A. R., & Taylor, J. P. (2010). Repeat expansion disease: Progress and puzzles in disease pathogenesis. *Nature Reviews. Genetics*, *11*(4), 247. https://doi.org/10.1038/NRG2748

Laço, M. N., Cortes, L., Travis, S. M., Paulson, H. L., & Rego, A. C. (2012). Valosin-containing protein (VCP/p97) is an activator of wild-type ataxin-3. *PloS One*, *7*(9). https://doi.org/10.1371/JOURNAL.PONE.0043563

Lahue, R. S., & Slater, D. L. (2003). DNA repair and trinucleotide repeat instability. *Frontiers in Bioscience : A Journal and Virtual Library*, *8*(SUPPL.). https://doi.org/10.2741/1107

Lam, Y. C., Bowman, A. B., Jafar-Nejad, P., Lim, J., Richman, R., Fryer, J. D., Hyun, E. D., Duvick, L. A., Orr, H. T., Botas, J., & Zoghbi, H. Y. (2006). ATAXIN-1 interacts with the repressor Capicua in its native complex to cause SCA1 neuropathology. *Cell*, *127*(7), 1335–1347. https://doi.org/10.1016/J.CELL.2006.11.038

Laskaratos, A., Breza, M., Karadima, G., & Koutsis, G. (2021). Wide range of reduced penetrance alleles in spinal and bulbar muscular atrophy: a model-based approach. *Journal of Medical Genetics*, *58*(6), 385–391. https://doi.org/10.1136/JMEDGENET-2020-106963

Lastres-Becker, I., Rüb, U., & Auburger, G. (2008). Spinocerebellar ataxia 2 (SCA2). *The Cerebellum 2008 7:2*, *7*(2), 115–124. https://doi.org/10.1007/S12311-008-0019-Y

Leavitt, B. R., Van Raamsdonk, J. M., Shehadeh, J., Fernandes, H., Murphy, Z., Graham, R. K., Wellington, C. L., Raymond, L. A., & Hayden, M. R. (2006). Wild-type huntingtin protects neurons from excitotoxicity. *J. Neurochem*, *96*, 1121–1129. https://doi.org/10.1111/j.1471-4159.2005.03605.x

Li, S. H., & Li, X. J. (1998). Aggregation of N-terminal huntingtin is dependent on the length of its glutamine repeats. *Human Molecular Genetics*, *7*(5), 777–782. https://doi.org/10.1093/HMG/7.5.777

Libby, R. T., Hagerman, K. A., Pineda, V. V., Lau, R., Cho, D. H., Baccam, S. L., Axford, M. M., Cleary, J. D., Moore, J. M., Sopher, B. L., Tapscott, S. J., Filippova, G. N., Pearson, C. E., & La Spada, A. R. (2008). CTCF cis-regulates trinucleotide repeat instability in an epigenetic manner: a novel basis for mutational hot spot determination. *PLoS Genetics*, *4*(11). https://doi.org/10.1371/JOURNAL.PGEN.1000257

Lieberman, A. P. (2018). Spinal and bulbar muscular atrophy. *Handbook of Clinical Neurology*, *148*, 625–632. https://doi.org/10.1016/B978-0-444-64076-5.00040-5

Lieberman, A. P., Shakkottai, V. G., & Albin, R. L. (2019). Polyglutamine Repeats in

Neurodegenerative Diseases. *Annual Review of Pathology*, *14*, 1. https://doi.org/10.1146/ANNUREV-PATHMECHDIS-012418-012857

Lima, L., & Coutinho, P. (1980). Clinical criteria for diagnosis of Machado-Joseph disease. *Neurology*, *30*(3), 319–319. https://doi.org/10.1212/WNL.30.3.319

Lin, Y., Dent, S. Y. R., Wilson, J. H., Wells, R. D., & Napierala, M. (2010). R loops stimulate genetic instability of CTG·CAG repeats. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(2), 692. https://doi.org/10.1073/PNAS.0909740107

Lin, Y., Hubert, L., & Wilson, J. H. (2009). TRANSCRIPTION DESTABILIZES TRIPLET REPEATS. *Molecular Carcinogenesis*, *48*(4), 350. https://doi.org/10.1002/MC.20488

Liu, G., Chen, X., Bissler, J. J., Sinden, R. R., & Leffak, M. (2010). Replication dependent instability at (CTG)•(CAG) repeat hairpins in human cells. *Nature Chemical Biology*, *6*(9), 652. https://doi.org/10.1038/NCHEMBIO.416

Liu, G., & Leffak, M. (2012). Instability of (CTG)n•(CAG)n trinucleotide repeats and DNA synthesis. *Cell & Bioscience 2012 2:1*, *2*(1), 1–13. https://doi.org/10.1186/2045-3701-2-7

Macdonald, M. E., Duyao, M., Calzonetti, T., Auerbach, A., Ryan, A., Barnes, G., White, J. K., Auerbach, W., Vonsattel, J. P., Gusella, J. F., & Joyner, A. L. (1996). Targeted inactivation of the mouse Huntington's disease gene homolog Hdh. *Cold Spring Harbor Symposia on Quantitative Biology*, *61*, 627–638. https://doi.org/10.1101/sqb.1996.061.01.063

Martindale, D., Hackam, A., Wieczorek, A., Ellerby, L., Wellington, C., McCutcheon, K., Singaraja, R., Kazemi-Esfarjani, P., Devon, R., Kim, S. U., Bredesen, D. E., Tufaro, F., & Hayden, M. R. (1998). Length of huntingtin and its polyglutamine tract influences localization and frequency of intracellular aggregates. *Nature Genetics*, *18*(2), 150–154. https://doi.org/10.1038/NG0298-150

Martinez-Vicente, M., Talloczy, Z., Wong, E., Tang, G., Koga, H., Kaushik, S., De Vries, R., Arias, E., Harris, S., Sulzer, D., & Cuervo, A. M. (2010). Cargo recognition failure is responsible for inefficient autophagy in Huntington's disease. *Nature Neuroscience*, *13*(5), 567–576. https://doi.org/10.1038/NN.2528

Maruyama, H., Izumi, Y., Morino, H., Oda, M., Hiromasa Toji, S. N., & Kawakami, H. (2002). Difference in disease-free survival curve and regional distribution according to subtype of spinocerebellar ataxia: a study of 1,286 Japanese patients. *American Journal of Medical Genetics*, *114*(5), 578–583. https://doi.org/10.1002/AJMG.10514

Matilla-Dueñas, A., Sánchez, I., Corral-Juan, M., Dávalos, A., Alvarez, R., & Latorre, P. (2009). Cellular and Molecular Pathways Triggering Neurodegeneration in the Spinocerebellar Ataxias. *The Cerebellum 2009 9:2*, *9*(2), 148–166. https://doi.org/10.1007/S12311-009-0144-2

McCampbell, A., Taylor, J. P., Taye, A. A., Robitschek, J., Li, M., Walcott, J., Merry, D., Chai, Y., Paulson, H., Sobue, G., & Fischbeck, K. H. (2000). CREB-binding protein sequestration by expanded polyglutamine. *Human Molecular Genetics*, *9*(14), 2197–2202. https://doi.org/10.1093/HMG/9.14.2197

McIntosh, C. S., Li, D., Wilton, S. D., & Aung-Htut, M. T. (2021). Polyglutamine Ataxias: Our Current Molecular Understanding and What the Future Holds for Antisense Therapies. *Biomedicines*, *9*(11). https://doi.org/10.3390/BIOMEDICINES9111499

McLoughlin, H. S., Moore, L. R., & Paulson, H. L. (2020). Pathogenesis of SCA3 and implications for other polyglutamine diseases. *Neurobiology of Disease*, *134*, 104635. https://doi.org/10.1016/J.NBD.2019.104635

McMurray, C. T. (2010). Mechanisms of trinucleotide repeat instability during human development. *Nature Reviews. Genetics*, *11*(11), 786. https://doi.org/10.1038/NRG2828

Meira, A. T., Pedroso, J. L., Boller, F., Franklin, G. L., Barsottini, O. G. P., & Teive, H. A. G. (2020). Reconstructing the History of Machado-Joseph Disease. *European Neurology*, *83*(1), 99–104. https://doi.org/10.1159/000507191

Miettinen, M. S., Monticelli, L., Nedumpully-Govindan, P., Knecht, V., & Ignatova, Z. (2014). Stable Polyglutamine Dimers Can Contain β-Hairpins with Interdigitated Side Chains—But Not α-Helices, β-Nanotubes, β-Pseudohelices, or Steric Zippers. *Biophysical Journal*, *106*(8), 1721. https://doi.org/10.1016/J.BPJ.2014.02.027

Miller, R. J. (1997). Calcium channels prove to be a real headache. *Trends in Neurosciences*, *20*(5), 189–192. https://doi.org/10.1016/S0166-2236(96)01037-5

Minakawa, E. N., & Nagai, Y. (2021). Protein Aggregation Inhibitors as Disease-Modifying Therapies for Polyglutamine Diseases. *Frontiers in Neuroscience*, *15*, 75. https://doi.org/10.3389/FNINS.2021.621996/BIBTEX

Mirkin, S. M. (2007). Expandable DNA repeats and human disease. *Nature*, *447*(7147), 932–940. https://doi.org/10.1038/NATURE05977

Mookerjee, S., Papanikolaou, T., Guyenet, S. J., Sampath, V., Lin, A., Vitelli, C., DeGiacomo, F., Sopher, B. L., Chen, S. F., La Spada, A. R., & Ellerby, L. M. (2009). Posttranslational modification of ataxin-7 at lysine 257 prevents autophagy-mediated turnover of an N-terminal caspase-7 cleavage fragment. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *29*(48), 15134–15144. https://doi.org/10.1523/JNEUROSCI.4720-09.2009

Morishima, Y., Wang, A. M., Yu, Z., Pratt, W. B., Osawa, Y., & Lieberman, A. P. (2008). CHIP deletion reveals functional redundancy of E3 ligases in promoting degradation of both signaling proteins and expanded glutamine proteins. *Human Molecular Genetics*, *17*(24), 3942–3952. https://doi.org/10.1093/HMG/DDN296

Murmann, A. E., Yu, J., Opal, P., & Peter, M. E. (2018). Trinucleotide Repeat Expansion Diseases, RNAi, and Cancer. *Trends in Cancer*, *4*(10), 684–700. https://doi.org/10.1016/J.TRECAN.2018.08.004

Mykowska, A., Sobczak, K., Wojciechowska, M., Kozlowski, P., & Krzyzosiak, W. J. (2011). CAG repeats mimic CUG repeats in the misregulation of alternative splicing. *Nucleic Acids Research*, *39*(20), 8938. https://doi.org/10.1093/NAR/GKR608

Naito, H., & Oyanagi, S. (1982). Familial myoclonus epilepsy and choreoathetosis: hereditary dentatorubral-pallidoluysian atrophy. *Neurology*, *32*(8), 798–807. https://doi.org/10.1212/WNL.32.8.798

Nakamori, M., Pearson, C. E., & Thornton, C. A. (2011). Bidirectional transcription stimulates expansion and contraction of expanded (CTG)•(CAG) repeats. *Human Molecular Genetics*, *20*(3), 580. https://doi.org/10.1093/HMG/DDQ501

Nakamura, Y., Tagawa, K., Oka, T., Sasabe, T., Ito, H., Shiwaku, H., La spada, A. R., & Okazawa, H. (2012). Ataxin-7 associates with microtubules and stabilizes the cytoskeletal network. *Human Molecular Genetics*, *21*(5), 1099–1110. https://doi.org/10.1093/HMG/DDR539

Nethisinghe, S., Pigazzini, M. L., Pemble, S., Sweeney, M. G., Labrum, R., Manso, K., Moore, D., Warner, J., Davis, M. B., & Giunti, P. (2018). PolyQ tract toxicity in SCA1 is length dependent in the absence of CAG repeat interruption. *Frontiers in Cellular Neuroscience*, *12*, 200. https://doi.org/10.3389/FNCEL.2018.00200/BIBTEX

Nóbrega, C., Matos, C., & Pereira de Almeida, L. (2018). Polyglutamine Disorders. *Journal of Neurochemistry*, *1049*(September), 155–162. https://doi.org/10.1007/978-3-319-71779-1

Nóbrega, C., & Pereira de Almeida, L. (Eds.). (2018). *Polyglutamine Disorders (SCA1)* (Vol. 1049). Springer International Publishing. https://doi.org/10.1007/978-3-319-71779-1

Nonhoff, U., Ralser, M., Welzel, F., Piccini, I., Balzereit, D., Yaspo, M. L., Lehrach, H., & Krobitsch, S. (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*(4), 1385–1396. https://doi.org/10.1091/MBC.E06-12-1120

Nonis, D., Schmidt, M. H. H., van de Loo, S., Eich, F., Dikic, I., Nowock, J., & Auburger, G. (2008). Ataxin-2 associates with the endocytosis complex and affects EGF receptor trafficking. *Cellular Signalling*, *20*(10), 1725–1739. https://doi.org/10.1016/J.CELLSIG.2008.05.018

Nucifora, J., Sasaki, M., Peters, M. F., Huang, H., Cooper, J. K., Yamada, M., Takahashi, H., Tsuji, S., Troncoso, J., Dawson, V. L., Dawson, T. M., & Ross, C. A. (2001). Interference by huntingtin and atrophin-1 with cbp-mediated transcription leading to cellular toxicity. *Science (New York, N.Y.)*, *291*(5512), 2423–2428. https://doi.org/10.1126/SCIENCE.1056784

Okazawa, H. (2003). Polyglutamine diseases: a transcription disorder? *Cellular and Molecular Life Sciences : CMLS*, *60*(7), 1427–1439. https://doi.org/10.1007/S00018-003-3013-Z

Olzscha, H., Schermann, S. M., Woerner, A. C., Pinkert, S., Hecht, M. H., Tartaglia, G. G., Vendruscolo, M., Hayer-Hartl, M., Hartl, F. U., & Vabulas, R. M. (2011). Amyloid-like aggregates sequester numerous metastable proteins with essential cellular functions. *Cell*, *144*(1), 67–78. https://doi.org/10.1016/J.CELL.2010.11.050

Opal, P., & Ashizawa, T. (2017a). Spinocerebellar Ataxia Type 1. *GeneReviews®*. https://www.ncbi.nlm.nih.gov/books/NBK1184/

Opal, P., & Orr, H. T. (2015). Animal Models of Spinocerebellar Ataxia Type 1. *Movement Disorders: Genetics and Models: Second Edition*, 979–990. https://doi.org/10.1016/B978-0-12-405195-9.00063-9

Ordway, J. M., Tallaksen-Greene, S., Gutekunst, C. A., Bernstein, E. M., Cearley, J. A., Wiener, H. W., Dure IV, L. S., Lindsey, R., Hersch, S. M., Jope, R. S., Albin, R. L., & Detloff, P. J. (1997). Ectopically Expressed CAG Repeats Cause Intranuclear Inclusions and a Progressive Late Onset Neurological Phenotype in the Mouse. *Cell*, *91*(6), 753–763. https://doi.org/10.1016/S0092-8674(00)80464-X

Orr, H. T. (2001). Beyond the Qs in the polyglutamine diseases. *Genes & Development*, *15*(8), 925–932. https://doi.org/10.1101/GAD.888401

Orr, Harry T. (2012). SCA1 – Phosphorylation, a Regulator of Atxain-1 Function and Pathogenesis. *Progress in Neurobiology*, *99*(3), 179. https://doi.org/10.1016/J.PNEUROBIO.2012.04.003

Orr, Harry T., & Zoghbi, H. Y. (2001). SCA1 molecular genetics: a history of a 13 year collaboration against glutamines. *Human Molecular Genetics*, *10*(20), 2307–2311. https://doi.org/10.1093/HMG/10.20.2307

Palhan, V. B., Chen, S., Peng, G. H., Tjernberg, A., Gamper, A. M., Fan, Y., Chait, B. T., La Spada, A. R., & Roeder, R. G. (2005). Polyglutamine-expanded ataxin-7 inhibits STAGA histone acetyltransferase activity to produce retinal degeneration. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(24), 8472–8477. https://doi.org/10.1073/PNAS.0503505102

Panov, A. V., Gutekunst, C. A., Leavitt, B. R., Hayden, M. R., Burke, J. R., Strittmatter, W. J., & Greenamyre, J. T. (2002). Early mitochondrial calcium defects in Huntington's disease are a direct effect of polyglutamines. *Nature Neuroscience*, *5*(8), 731–736. https://doi.org/10.1038/NN884

Paul, S., Dansithong, W., Figueroa, K. P., Scoles, D. R., & Pulst, S. M. (2018). Staufen1 links RNA stress granules and autophagy in a model of neurodegeneration. *Nature Communications*, *9*(1). https://doi.org/10.1038/S41467-018-06041-3

Paulson, H. L., Perez, M. K., Trottier, Y., Trojanowski, J. Q., Subramony, S. H., Das, S. S., Vig, P., Mandel, J. L., Fischbeck, K. H., & Pittman, R. N. (1997). Intranuclear Inclusions of Expanded Polyglutamine Protein in Spinocerebellar Ataxia Type 3. *Neuron*, *19*(2), 333–344. https://doi.org/10.1016/S0896-6273(00)80943-5

Paulson, H. L., & Shakkottai, V. (2020). Spinocerebellar Ataxia Type 3. *GeneReviews®*. https://www.ncbi.nlm.nih.gov/books/NBK1196/

Paulson, H. L. (2009). The Spinocerebellar Ataxias. *Journal of Neuro-Ophthalmology : The Official Journal of the North American Neuro-Ophthalmology Society*, *29*(3), 227. https://doi.org/10.1097/WNO0B013E3181B416DE

Paulson, H. L., Shakkottai, V. G., Clark, H. B., & Orr, H. T. (2017). Polyglutamine spinocerebellar ataxias — from genes to potential treatments. *Nature Reviews. Neuroscience*, *18*(10), 613. https://doi.org/10.1038/NRN.2017.92

Pearson, C. E., Edamura, K. N., & Cleary, J. D. (2005). Repeat instability: mechanisms of dynamic mutations. *Nature Reviews. Genetics*, *6*(10), 729–742. https://doi.org/10.1038/NRG1689

Pellistri, F., Bucciantini, M., Invernizzi, G., Gatta, E., Penco, A., Frana, A. M., Nosi, D., Relini, A., Regonesi, M. E., Gliozzi, A., Tortora, P., Robello, M., & Stefani, M. (2013). Different ataxin-3 amyloid aggregates induce intracellular Ca(2+)

deregulation by different mechanisms in cerebellar granule cells. *Biochimica et Biophysica Acta*, *1833*(12), 3155–3165. https://doi.org/10.1016/J.BBAMCR.2013.08.019

Perutz, M. F., Johnson, T., Suzuki, M., & Finch, J. T. (1994). Glutamine repeats as polar zippers: their possible role in inherited neurodegenerative diseases. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(12), 5355. https://doi.org/10.1073/PNAS.91.12.5355

Peterson, M. G., Tanese, N., Franklin Pugh, B., & Tjian, R. (1990). Functional domains and upstream activation properties of cloned human TATA binding protein. *Science (New York, N.Y.)*, *248*(4963), 1625–1630. https://doi.org/10.1126/SCIENCE.2363050

Podvin, S., Reardon, H. T., Yin, K., Mosier, C., & Hook, V. (2019). Multiple clinical features of Huntington's disease correlate with mutant HTT gene CAG repeat lengths and neurodegeneration. *Journal of Neurology*, *266*(3), 551–564. https://doi.org/10.1007/S00415-018-8940-6

Potaman, V. N., & Sinden, R. R. (2013). *DNA: Alternative Conformations and Biology*. https://www.ncbi.nlm.nih.gov/books/NBK6545/

Pringsheim, T., Wiltshire, K., Day, L., Dykeman, J., Steeves, T., & Jette, N. (2012). The incidence and prevalence of Huntington's disease: a systematic review and meta-analysis. *Movement Disorders : Official Journal of the Movement Disorder Society*, *27*(9), 1083–1091. https://doi.org/10.1002/MDS.25075

Pulst, S. M. (2019). Spinocerebellar Ataxia Type 2. *GeneReviews®*. https://www.ncbi.nlm.nih.gov/books/NBK1275/

Qi, L., Zhang, X. D., Wu, J. C., Lin, F., Wang, J., DiFiglia, M., & Qin, Z. H. (2012). The role of chaperone-mediated autophagy in huntingtin degradation. *PloS One*, *7*(10). https://doi.org/10.1371/JOURNAL.PONE.0046834

Ralser, M., Albrecht, M., Nonhoff, U., Lengauer, T., Lehrach, H., & Krobitsch, S. (2005). An integrative approach to gain insights into the cellular function of human ataxin-2. *Journal of Molecular Biology*, *346*(1), 203–214. https://doi.org/10.1016/J.JMB.2004.11.024

Ravikumar, B., Vacher, C., Berger, Z., Davies, J. E., Luo, S., Oroz, L. G., Scaravilli, F., Easton, D. F., Duden, R., O'Kane, C. J., & Rubinsztein, D. C. (2004). Inhibition of mTOR induces autophagy and reduces toxicity of polyglutamine expansions in fly and mouse models of Huntington disease. *Nature Genetics*, *36*(6), 585–595. https://doi.org/10.1038/NG1362

Rawlins, M. D., Wexler, N. S., Wexler, A. R., Tabrizi, S. J., Douglas, I., Evans, S. J. W., & Smeeth, L. (2016). The Prevalence of Huntington's Disease. *Neuroepidemiology*, *46*(2), 144–153. https://doi.org/10.1159/000443738

Raymond, L. A., André, V. M., Cepeda, C., Gladding, C. M., Milnerwood, A. J., & Levine, M. S. (2011). Pathophysiology of Huntington's Disease: Time-Dependent Alterations in Synaptic and Receptor Function. *Neuroscience*, *198*, 252. https://doi.org/10.1016/J.NEUROSCIENCE.2011.08.052

Raymond, Lynn A. (2017). Striatal synaptic dysfunction and altered calcium regulation

in Huntington disease. *Biochemical and Biophysical Research Communications*, *483*(4), 1051–1062. https://doi.org/10.1016/J.BBRC.2016.07.058

Rego, A. C., & de Almeida, L. P. (2005). Molecular targets and therapeutic strategies in Huntington's Disease. *Current Drug Targets: CNS and Neurological Disorders*, *4*(4), 361–381. https://doi.org/10.2174/1568007054546081

Reiner, A., Dragatsis, I., & Dietrich, P. (2011). GENETICS AND NEUROPATHOLOGY OF HUNTINGTON'S DISEASE. *International Review of Neurobiology*, *98*, 325. https://doi.org/10.1016/B978-0-12-381328-2.00014-6

Rigamonti, D., Bauer, J. H., De-Fraja, C., Conti, L., Sipione, S., Sciorati, C., Clementi, E., Hackam, A., Hayden, M. R., Li, Y., Cooper, J. K., Ross, C. A., Govoni, S., Vincenz, C., & Cattaneo, E. (2000). Wild-Type Huntingtin Protects from Apoptosis Upstream of Caspase-3. *The Journal of Neuroscience*, *20*(10), 3705. https://doi.org/10.1523/JNEUROSCI.20-10-03705.2000

Rigamonti, D., Sipione, S., Goffredo, D., Zuccato, C., Fossale, E., & Cattaneo, E. (2001). Huntingtin's Neuroprotective Activity Occurs via Inhibition of Procaspase-9 Processing *. *Journal of Biological Chemistry*, *276*(18), 14545–14548. https://doi.org/10.1074/JBC.C100044200

Riley, B. E., Zoghbi, H. Y., & Orr, H. T. (2005). SUMOylation of the Polyglutamine Repeat Protein, Ataxin-1, Is Dependent on a Functional Nuclear Localization Signal. *Journal of Biological Chemistry*, *280*(23), 21942–21948. https://doi.org/10.1074/JBC.M501677200

Robertson, A. L., Bate, M. A., Androulakis, S. G., Bottomley, S. P., & Buckle, A. M. (2011). PolyQ: a database describing the sequence and domain context of polyglutamine repeats in proteins. *Nucleic Acids Research*, *39*(suppl_1), D272–D276. https://doi.org/10.1093/NAR/GKQ1100

Rodrigues, A. J., do Carmo Costa, M., Silva, T. L., Ferreira, D., Bajanca, F., Logarinho, E., & Maciel, P. (2010). Absence of ataxin-3 leads to cytoskeletal disorganization and increased cell death. *Biochimica et Biophysica Acta*, *1803*(10), 1154–1163. https://doi.org/10.1016/J.BBAMCR.2010.07.004

Rosas, H. D., Tuch, D. S., Hevelone, N. D., Zaleta, A. K., Vangel, M., Hersch, S. M., & Salat, D. H. (2006). Diffusion tensor imaging in presymptomatic and early Huntington's disease: Selective white matter pathology and its relationship to clinical measures. *Movement Disorders*, *21*(9), 1317–1325. https://doi.org/10.1002/MDS.20979

Ross, C. A. (1997). Intranuclear Neuronal Inclusions: A Common Pathogenic Mechanism for Glutamine-Repeat Neurodegenerative Diseases? *Neuron*, *19*(6), 1147–1150. https://doi.org/10.1016/S0896-6273(00)80405-5

Ross, C. A., & Poirier, M. A. (2004). Protein aggregation and neurodegenerative disease. *Nature Medicine*, *10 Suppl*(7), S10. https://doi.org/10.1038/NM1066

Rousseaux, M. W. C., Tschumperlin, T., Lu, H. C., Lackey, E. P., Bondar, V. V., Wan, Y. W., Tan, Q., Adamski, C. J., Friedrich, J., Twaroski, K., Chen, W., Tolar, J., Henzler, C., Sharma, A., Bajić, A., Lin, T., Duvick, L., Liu, Z., Sillitoe, R. V., … Orr, H. T. (2018). ATXN1-CIC Complex Is the Primary Driver of Cerebellar Pathology in Spinocerebellar Ataxia Type 1 through a Gain-of-Function

Mechanism. *Neuron*, *97*(6), 1235. https://doi.org/10.1016/J.NEURON.2018.02.013

Roy, D., & Lieber, M. R. (2009). G Clustering Is Important for the Initiation of Transcription-Induced R-Loops In Vitro, whereas High G Density without Clustering Is Sufficient Thereafter. *Molecular and Cellular Biology*, *29*(11), 3124. https://doi.org/10.1128/MCB.00139-09

Rüb, U., Bürk, K., Timmann, D., den Dunnen, W., Seidel, K., Farrag, K., Brunt, E., Heinsen, H., Egensperger, R., Bornemann, A., Schwarzacher, S., Korf, H. W., Schöls, L., Bohl, J., & Deller, T. (2012). Spinocerebellar ataxia type 1 (SCA1): new pathoanatomical and clinico-pathological insights. *Neuropathology and Applied Neurobiology*, *38*(7), 665–680. https://doi.org/10.1111/J.1365-2990.2012.01259.X

Saegusa, H., Wakamori, M., Matsuda, Y., Wang, J., Mori, Y., Zong, S., & Tanabe, T. (2007). Properties of human Cav2.1 channel with a spinocerebellar ataxia type 6 mutation expressed in Purkinje cells. *Molecular and Cellular Neurosciences*, *34*(2), 261–270. https://doi.org/10.1016/J.MCN.2006.11.006

Satterfield, T. F., & Pallanck, L. J. (2006). Ataxin-2 and its Drosophila homolog, ATX2, physically assemble with polyribosomes. *Human Molecular Genetics*, *15*(16), 2523–2532. https://doi.org/10.1093/HMG/DDL173

Saudou, F., & Humbert, S. (2016). The Biology of Huntingtin. *Neuron*, *89*(5), 910–926. https://doi.org/10.1016/J.NEURON.2016.02.003

Schöls, L., Vieira-saecker, A. M. M., Schöls, S., Przuntek, H., Epplen, J. T., & Riess, O. (1995). Trinucleotide expansion within the MJD1 gene presents clinically as spinocerebellar ataxia and occurs most frequently in German SCA patients. *Human Molecular Genetics*, *4*(6), 1001–1005. https://doi.org/10.1093/HMG/4.6.1001

Sehn, J. K. (2015). Insertions and Deletions (Indels). *Clinical Genomics*, 129–150. https://doi.org/10.1016/B978-0-12-404748-8.00009-5

Seidel, K., Meister, M., Dugbartey, G. J., Zijlstra, M. P., Vinet, J., Brunt, E. R. P., Van Leeuwen, F. W., Rüb, U., Kampinga, H. H., & Den Dunnen, W. F. A. (2012). Cellular protein quality control and the evolution of aggregates in spinocerebellar ataxia type 3 (SCA3). *Neuropathology and Applied Neurobiology*, *38*(6), 548–558. https://doi.org/10.1111/J.1365-2990.2011.01220.X

Servadio, A., Koshy, B., Armstrong, D., Antalffy, B., Orr, H. T., & Zoghbi, H. Y. (1995). Expression analysis of the ataxin-1 protein in tissues from normal and spinocerebellar ataxia type 1 individuals. *Nature Genetics*, *10*(1), 94–98. https://doi.org/10.1038/NG0595-94

Shao, J., & Diamond, M. I. (2007). Polyglutamine diseases: emerging concepts in pathogenesis and therapy. *Human Molecular Genetics*, *16*(R2), R115–R123. https://doi.org/10.1093/HMG/DDM213

Sharp, A. H., Loev, S. J., Schilling, G., Li, S. H., Li, X. J., Bao, J., Wagster, M. V., Kotzuk, J. A., Steiner, J. P., Lo, A., Hedreen, J., Sisodia, S., Snyder, S. H., Dawson, T. M., Ryugo, D. K., & Ross, C. A. (1995). Widespread expression of Huntington's disease gene (IT15) protein product. *Neuron*, *14*(5), 1065–1074. https://doi.org/10.1016/0896-6273(95)90345-3

Shimohata, M., Shimohata, T., Igarashi, S., Naruse, S., & Tsuji, S. (2005). Interference

of CREB-dependent transcriptional activation by expanded polyglutamine stretches – augmentation of transcriptional activation as a potential therapeutic strategy for polyglutamine diseases. *Journal of Neurochemistry*, *93*(3), 654–663. https://doi.org/10.1111/J.1471-4159.2005.03060.X

Singh, B. K., Vatsa, N., Nelson, V. K., Kumar, V., Kumar, S. S., Mandal, S. C., Pal, M., & Jana, N. R. (2018). Azadiradione Restores Protein Quality Control and Ameliorates the Disease Pathogenesis in a Mouse Model of Huntington's Disease. *Molecular Neurobiology*, *55*(8), 6337–6346. https://doi.org/10.1007/S12035-017-0853-3

Smith, J. K., Gonda, V. E., & Malamud, N. (1958). Unusual form of cerebellar ataxia. *Neurology*, *8*(3), 205–205. https://doi.org/10.1212/WNL.8.3.205

Smith, R., Brundin, P., & Li, J. Y. (2005). Synaptic dysfunction in Huntington's disease: a new perspective. *Cellular and Molecular Life Sciences : CMLS*, *62*(17), 1901–1912. https://doi.org/10.1007/S00018-005-5084-5

Stoyas, C. A., & La Spada, A. R. (2018). The CAG-polyglutamine repeat diseases: a clinical, molecular, genetic, and pathophysiologic nosology. *Handbook of Clinical Neurology*, *147*, 143–170. https://doi.org/10.1016/B978-0-444-63233-3.00011-7

Sugimoto, N., Nakano, S. ichi, Katoh, M., Matsumura, A., Nakamuta, H., Ohmichi, T., Yoneyama, M., & Sasaki, M. (1995). Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes. *Biochemistry*, *34*(35), 11211–11216. https://doi.org/10.1021/BI00035A029

Sullivan, R., Yau, W. Y., O'Connor, E., & Houlden, H. (2018). Spinocerebellar ataxia: an update. *Journal of Neurology 2018 266:2*, *266*(2), 533–544. https://doi.org/10.1007/S00415-018-9076-4

Szlachcic, W. J., Switonski, P. M., Kurkowiak, M., Wiatr, K., & Figiel, M. (2015). Mouse polyQ database: A new online resource for research using mouse models of neurodegenerative diseases. *Molecular Brain*, *8*(1), 1–8. https://doi.org/10.1186/S13041-015-0160-8/FIGURES/3

Takahashi, A. (2001). Hiroshi Kawahara (1858-1918). *Journal of Neurology*, *248*(3), 241–242. https://doi.org/10.1007/S004150170236

Takahashi, T., Katada, S., & Onodera, O. (2010). Polyglutamine Diseases: Where does Toxicity Come from? What is Toxicity? Where are We Going? *Journal of Molecular Cell Biology*, *2*(4), 180–191. https://doi.org/10.1093/JMCB/MJQ005

Takeuchi, T., & Nagai, Y. (2017). Protein Misfolding and Aggregation as a Therapeutic Target for Polyglutamine Diseases. *Brain Sciences*, *7*(10). https://doi.org/10.3390/BRAINSCI7100128

Tanaka, F., Doyu, M., Ito, Y., Matsumoto, M., Mitsuma, T., Abe, K., Aoki, M., Itoyama, Y., Fischbeck, K. H., & Sobue, G. (1996). Founder Effect in Spinal and Bulbar Muscular Atrophy (SBMA). *Human Molecular Genetics*, *5*(9), 1253–1257. https://doi.org/10.1093/HMG/5.9.1253

Tang, B., Liu, C., Shen, L., Dai, H., Pan, Q., Jing, L., Ouyang, S., & Xia, J. (2000). Frequency of SCA1, SCA2, SCA3/MJD, SCA6, SCA7, and DRPLA CAG trinucleotide repeat expansion in patients with hereditary spinocerebellar ataxia from

Chinese kindreds. *Archives of Neurology*, *57*(4), 540–544. https://doi.org/10.1001/archneur.57.4.540

Tao, T., & Tartakoff, A. M. (2001). Nuclear relocation of normal huntingtin. *Traffic (Copenhagen, Denmark)*, *2*(6), 385–394. https://doi.org/10.1034/J.1600-0854.2001.002006385.X

The web framework for perfectionists with deadlines | Django. (n.d.). Retrieved August 19, 2022, from https://www.djangoproject.com/

Titica, J., & Van Bogaert, L. (1946). HEREDO-DEGENERATIVE HEMIBALLISMUSA Contribution to the Question of Primary Atrophy of the Corpus Luysii. *Brain*, *69*(4), 251–263. https://doi.org/10.1093/BRAIN/69.4.251

Todd, T. W., & Lim, J. (2013). Aggregation Formation in the Polyglutamine Diseases: Protection at a Cost? *Molecules and Cells*, *36*(3), 185. https://doi.org/10.1007/S10059-013-0167-X

Toyoshima, Y., Onodera, O., Yamada, M., Tsuji, S., & Takahashi, H. (2022). Spinocerebellar Ataxia Type 17. *GeneReviews®*. https://www.ncbi.nlm.nih.gov/books/NBK1438/

Toyoshima, Y., & Takahashi, H. (2018). Spinocerebellar Ataxia Type 17 (SCA17). *Advances in Experimental Medicine and Biology*, *1049*, 219–231. https://doi.org/10.1007/978-3-319-71779-1_10

Trushina, E., & McMurray, C. T. (2007). Oxidative stress and mitochondrial dysfunction in neurodegenerative diseases. *Neuroscience*, *145*(4), 1233–1248. https://doi.org/10.1016/J.NEUROSCIENCE.2006.10.056

Tsai, H. F., Tsai, H. J., & Hsieh, M. (2004). Full-length expanded ataxin-3 enhances mitochondrial-mediated cell death and decreases Bcl-2 expression in human neuroblastoma cells. *Biochemical and Biophysical Research Communications*, *324*(4), 1274–1282. https://doi.org/10.1016/J.BBRC.2004.09.192

Tsuji, S. (2011). Chapter 41 - Dentatorubral-pallidoluysian atrophy. *Ataxic Disorders Vol. 103*, *103*, 587–594. https://doi.org/10.1016/B978-0-444-51892-7.00041-3

Tsuji, S., Onodera, O., Goto, J., & Nishizawa, M. (2008). Sporadic ataxias in Japan--a population-based epidemiological study. *Cerebellum (London, England)*, *7*(2), 189–197. https://doi.org/10.1007/S12311-008-0028-X

Tunalı, N. E. (2020). Molecular Mechanisms of Polyglutamine Pathology and Lessons Learned from Huntington's Disease. *Neurodegenerative Diseases - Molecular Mechanisms and Current Therapeutic Approaches*. https://doi.org/10.5772/INTECHOPEN.93508

Types of Database Management Systems. (n.d.). Retrieved July 26, 2022, from https://www.c-sharpcorner.com/UploadFile/65fc13/types-of-database-management-systems/

Usage Statistics of JavaScript as Client-side Programming Language on Websites, August 2022. (n.d.). Retrieved August 19, 2022, from https://w3techs.com/technologies/details/cp-javascript/

Van De Warrenburg, B. P. C., Frenken, C. W. G. M., Ausems, M. G. E. M., Kleefstra,

T., Sinke, R. J., Knoers, N. V. A. M., & Kremer, H. P. H. (2001). Striking anticipation in spinocerebellar ataxia type 7: the infantile phenotype. *Journal of Neurology*, *248*(10), 911–914. https://doi.org/10.1007/S004150170082

Van den Broek, W. J. A. A., Wansink, D. G., & Wieringa, B. (2007). Somatic CTG•CAG repeat instability in a mouse model for myotonic dystrophy type 1 is associated with changes in cell nuclearity and DNA ploidy. *BMC Molecular Biology*, *8*, 61. https://doi.org/10.1186/1471-2199-8-61

Velázquez-Pérez, L., Rodríguez-Labrada, R., García-Rodríguez, J. C., Almaguer-Mederos, L. E., Cruz-Mariño, T., & Laffita-Mesa, J. M. (2011). A Comprehensive Review of Spinocerebellar Ataxia Type 2 in Cuba. *The Cerebellum 2011 10:2*, *10*(2), 184–198. https://doi.org/10.1007/S12311-011-0265-2

Veneziano, L., & Frontali, M. (2016). DRPLA. *Encyclopedia of Genetics, Genomics, Proteomics and Informatics*, 563–563. https://doi.org/10.1007/978-1-4020-6754-9_4901

Wadia, N. H., & Swami, R. K. (1971). A new form of heredo-familial spinocerebellar degeneration with slow eye movements (nine families). *Brain : A Journal of Neurology*, *94*(2), 359–374. https://doi.org/10.1093/BRAIN/94.2.359

Wang, G. H., Sawai, N., Kotliarova, S., Kanazawa, I., & Nukina, N. (2000). Ataxin-3, the MJD1 gene product, interacts with the two human homologs of yeast DNA repair protein RAD23, HHR23A and HHR23B. *Human Molecular Genetics*, *9*(12), 1795–1803. https://doi.org/10.1093/HMG/9.12.1795

Wang, G., & Vasquez, K. M. (2006). Non-B DNA structure-induced genetic instability. *Mutation Research*, *598*(1–2), 103–119. https://doi.org/10.1016/J.MRFMMM.2006.01.019

Wang, Q., Li, A., & Ye, Y. (2006). Regulation of retrotranslocation by p97-associated deubiquitinating enzyme ataxin-3. *The Journal of Cell Biology*, *174*(7), 963–971. https://doi.org/10.1083/JCB.200605100

Watanabe, R., Higashi, S., Nonaka, T., Kawakami, I., Oshima, K., Niizato, K., Akiyama, H., Yoshida, M., Hasegawa, M., & Arai, T. (2020). Intracellular dynamics of Ataxin-2 in the human brains with normal and frontotemporal lobar degeneration with TDP-43 inclusions. *Acta Neuropathologica Communications*, *8*(1). https://doi.org/10.1186/S40478-020-01055-9

Watase, K., Barrett, C. F., Miyazaki, T., Ishiguro, T., Ishikawa, K., Hu, Y., Unno, T., Sun, Y., Kasai, S., Watanabe, M., Gomez, C. M., Mizusawa, H., Tsien, R. W., & Zoghbi, H. Y. (2008). Spinocerebellar ataxia type 6 knockin mice develop a progressive neuronal dysfunction with age-dependent accumulation of mutant CaV2.1 channels. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(33), 11987–11992. https://doi.org/10.1073/PNAS.0804350105/SUPPL_FILE/0804350105SI.PDF

Wellington, C. L., Singaraja, R., Ellerby, L., Savill, J., Roy, S., Leavitt, B., Cattaneo, E., Hackam, A., Sharp, A., Thornberry, N., Nicholson, D. W., Bredesen, D. E., & Hayden, M. R. (2000). Inhibiting caspase cleavage of huntingtin reduces toxicity and aggregate formation in neuronal and nonneuronal cells. *The Journal of Biological Chemistry*, *275*(26), 19831–19838.

https://doi.org/10.1074/JBC.M001475200

What Is a Database | Oracle Portugal. (n.d.). Retrieved July 26, 2022, from https://www.oracle.com/pt/database/what-is-database/

Wissmann, M., Yin, N., Müller, J. M., Greschik, H., Günther, T., Metzger, E., Schüle, R., Fodor, B. D., Jenuwein, T., Vogler, C., Schneider, R., & Buettner, R. (2007). Cooperative demethylation by JMJD2C and LSD1 promotes androgen receptor-dependent gene expression. *Nature Cell Biology*, *9*(3), 347–353. https://doi.org/10.1038/NCB1546

Yazawa, I., Nukina, N., Hashida, H., Goto, J., Yamada, M., & Kanazawa, I. (1995). Abnormal gene product identified in hereditary dentatorubral-pallidoluysian atrophy (DRPLA) brain. *Nature Genetics*, *10*(1), 99–103. https://doi.org/10.1038/NG0595-99

Yu, Y. C., Kuo, C. L., Cheng, W. L., Liu, C. S., & Hsieh, M. (2009). Decreased antioxidant enzyme activity and increased mitochondrial DNA damage in cellular models of Machado-Joseph disease. *Journal of Neuroscience Research*, *87*(8), 1884–1891. https://doi.org/10.1002/JNR.22011

Yue, S., Serra, H. G., Zoghbi, H. Y., & Orr, H. T. (2001). The spinocerebellar ataxia type 1 protein, ataxin-1, has RNA-binding activity that is inversely affected by the length of its polyglutamine tract. *Human Molecular Genetics*, *10*(1), 25–30. https://doi.org/10.1093/HMG/10.1.25

Zhang, F., Xu, D., Yuan, L., Sun, Y., & Xu, Z. (2014). Epigenetic regulation of Atrophin1 by lysine-specific demethylase 1 is required for cortical progenitor maintenance. *Nature Communications*, *5*. https://doi.org/10.1038/NCOMMS6815

Zhang, Y., Li, M., Drozda, M., Chen, M., Ren, S., Sanchez, R. O. M., Leavitt, B. R., Cattaneo, E., Ferrante, R. J., Hayden, M. R., & Friedlander, R. M. (2003). Depletion of wild-type huntingtin in mouse models of neurologic diseases. *Journal of Neurochemistry*, *87*(1), 101–106. https://doi.org/10.1046/J.1471-4159.2003.01980.X

Zhao, J., Bacolla, A., Wang, G., & Vasquez, K. M. (2010). Non-B DNA structure-induced genetic instability and evolution. *Cellular and Molecular Life Sciences : CMLS*, *67*(1), 43. https://doi.org/10.1007/S00018-009-0131-2

Zhao, X. N., & Usdin, K. (2015). The Repeat Expansion Diseases: the dark side of DNA repair? *DNA Repair*, *32*, 96. https://doi.org/10.1016/J.DNAREP.2015.04.019

Zhivotovsky, B., & Orrenius, S. (2011). Calcium and cell death mechanisms: a perspective from the cell death community. *Cell Calcium*, *50*(3), 211–221. https://doi.org/10.1016/J.CECA.2011.03.003

Zhuchenko, O., Bailey, J., Bonnen, P., Ashizawa, T., Stockton, D. W., Amos, C., Dobyns, W. B., Subramony, S. H., Zoghbi, H. Y., & Lee, C. C. (1997). Autosomal dominant cerebellar ataxia (SCA6) associated with small polyglutamine expansions in the alpha 1A-voltage-dependent calcium channel. *Nature Genetics*, *15*(1), 62–69. https://doi.org/10.1038/NG0197-62

Zoghbi, H. Y., & Orr, H. T. (2009). Pathogenic mechanisms of a polyglutamine-mediated neurodegenerative disease, spinocerebellar ataxia type 1. *The Journal of Biological*

Chemistry, *284*(12), 7425–7429. https://doi.org/10.1074/JBC.R800041200

Zuccato, C., Ciammola, A., Rigamonti, D., Leavitt, B. R., Goffredo, D., Conti, L., MacDonald, M. E., Friedlander, R. M., Silani, V., Hayden, M. R., Timmusk, T., Sipione, S., & Cattaneo, E. (2001). Loss of huntingtin-mediated BDNF gene transcription in Huntington's disease. *Science*, *293*(5529), 493–498. https://doi.org/10.1126/SCIENCE.1059581/SUPPL_FILE/1059581S3_THUMB.GIF

Zuccato, Chiara, Belyaev, N., Conforti, P., Ooi, L., Tartari, M., Papadimou, E., MacDonald, M., Fossale, E., Zeitlin, S., Buckley, N., & Cattaneo, E. (2007). Widespread Disruption of Repressor Element-1 Silencing Transcription Factor/Neuron-Restrictive Silencer Factor Occupancy at Its Target Genes in Huntington's Disease. *The Journal of Neuroscience*, *27*(26), 6972. https://doi.org/10.1523/JNEUROSCI.4278-06.2007

Zuccato, Chiara, & Cattaneo, E. (2009). Brain-derived neurotrophic factor in neurodegenerative diseases. *Nature Reviews. Neurology*, *5*(6), 311–322. https://doi.org/10.1038/NRNEUROL.2009.54

Zuccato, Chiara, Tartari, M., Crotti, A., Goffredo, D., Valenza, M., Conti, L., Cataudella, T., Leavitt, B. R., Hayden, M. R., Timmusk, T., Rigamonti, D., & Cattaneo, E. (2003). Huntingtin interacts with REST/NRSF to modulate the transcription of NRSE-controlled neuronal genes. *Nature Genetics*, *35*(1), 76–83. https://doi.org/10.1038/NG1219