EXPANSION OF TRINUCLEOTIDE REPEATS IN HUMAN DNA AND MOLECULAR DIAGNOSIS OF TRINUCLEOTIDE EXPANSION DISORDERS: INTEGRATIVE REVIEW

EXPANSÃO DE REPETIÇÕES DE TRINUCLEOTÍDEOS NO DNA HUMANO E DIAGNÓSTICO MOLECULAR DE TRANSTORNOS DE EXPANSÃO DE TRINUCLEOTÍDEOS: REVISÃO INTEGRATIVA

Leoncio Rabelo Borges Filho, André Luiz Pedrosa

Departamento de Bioquímica, Farmacologia e Fisiologia, ICBN, Universidade Federal do Triângulo Mineiro, andre.pedrosa@uftm.edu.br

ABSTRACT

Trinucleotide repeat (TNR) expansions are increases in the number of repeated trinucleotides in the genome. TNR expansions have been confirmed as the molecular etiology of various neurodegenerative disorders, such as Huntington's Disease, Fragile X Syndrome, and Friedreich's Ataxia. Due to the potential achieved by molecular diagnostic technology, new disorders related to TNR expansions are being elucidated more rapidly, as high-throughput technologies for analyzing DNA are described every year. Thus, we carried out an integrative review on the importance of improving DNA analysis methods for the molecular diagnosis of diseases related to TNR expansions. We searched the PubMed using the descriptors "trinucleotide expansion", "DNA analysis", "genome" and "DNA sequencing". In our review, 42 genomic targets were identified as being responsible for diseases related to TNR expansions, with one target being related to two diseases and two targets being related to three diseases. The rest of the targets are related to a single disease. Despite the emergence of new sequencing technologies, PCR-based techniques continue to play an important role in identifying targets related to diseases caused by TNR expansions. New DNA sequencing technologies have demonstrated an important contribution to the identification of targets related to diseases caused by TNR, especially since 2019. The development of novel DNA analysis techniques and new genomic analysis tools has the potential to unveil novel targets associated with TNR expansions.

KEYWORDS: Trinucleotide expansions. DNA sequencing. Molecular Biology. Molecular Diagnostics. Genetics.

RESUMO

As expansões de repetições de trinucleotídeos (TNR) são aumentos no número de trinucleotídeos repetidos no genoma. As expansões TNR foram confirmadas como a etiologia molecular de várias doenças neurodegenerativas, como a doença de Huntington, a síndrome do X frágil e a ataxia de Friedreich. Devido ao potencial alcançado pela tecnologia de diagnóstico molecular, novas doenças relacionadas com expansões de TNR estão a ser elucidadas mais rapidamente, uma vez que são descritas todos os anos tecnologias de alto rendimento para analisar o DNA. Assim,

realizamos uma revisão integrativa sobre a importância do aprimoramento dos métodos de análise de DNA para o diagnóstico molecular de doenças relacionadas às expansões de TNR. Pesquisámos na PubMed utilizando os descritores “trinucleotide expansion”, “DNA analysis”, “genome” e “DNA sequencing”. Na nossa revisão, foram identificados 42 alvos genômicos como responsáveis por doenças relacionadas com expansões de TNR, sendo que um alvo está relacionado com duas doenças e dois alvos estão relacionados com três doenças. Os restantes alvos estão relacionados com uma única doença. Apesar do aparecimento de novas tecnologias de sequenciamento, as técnicas baseadas na PCR continuam a desempenhar um papel importante na identificação de alvos relacionados com doenças causadas por expansões de TNR. Nos últimos anos, as novas tecnologias de sequenciamento de DNA têm demonstrado contribuição importante para a identificação de alvos relacionados com doenças causadas por TNR.
PALAVRAS-CHAVE: Expansões de trinucleotídeos. Sequenciamento de DNA. Biologia molecular. Diagnóstico molecular. Genética.

INTRODUCTION

The human genome corresponds to the information present in the 23 pairs of nuclear chromosomes, and more than 50% of this genetic material corresponds to repeated sequences[1]. Depending on how they appear on the chromosome, these repeats can be classified as "tandem" or "dispersed". Tandem repeats have continuous clusters of common nucleotides. Dispersed repeats correspond to sequence segments that are scattered in different regions of a genome. Within the tandem repeat category, the group of repeated nucleotides can be classified as microsatellites and minisatellites, depending on the number of nucleotides that make up the repeat unit. Microsatellites and minisatellites are repeats made up of 1-10 nucleotides and 11-100 nucleotides, respectively[2]. Trinucleotide repeats (TNR) are among the most common in the human genome and it is noteworthy in the literature that trinucleotide repeats may compose unstable parts of the genome. Due to their instability, these repeat regions can undergo expansions during the DNA replication process. These expansions have pathological consequences for an individual and for his/her descendants. In addition, the greater the number of TNR, the earlier the onset of the disorder related to this expansion, this characteristic being identified as "anticipation"[3]. Despite this, affected individuals with the same

number of CAG repeats in the ATXN3 gene can have a markedly different age of onset of the disease[4], demonstrating other factors that interfere in the pathogenesis of TNR diseases.

Disorders caused by TNR expansions are mainly neurodegenerative in nature. In other words, these conditions primarily cause the progressive loss of function in nerve cells, leading to neurological symptoms. Related disorders include Fragile X Syndrome, Myotonic Dystrophy, Huntington's Disease, Friedreich's Ataxia, Spinocerebellar Ataxia, Fragile X-Associated Ataxia-Tremor Syndrome, among others[5]. Disorders related to abnormal polyglutamine repeats (PolyQ), i.e. disorders in which proteins form with abnormal amounts of glutamine repeats in their primary structure, are caused by expansions of the CAG repeat. No neurodegenerative ataxia has been categorized as a polyQ disorder in the last two decades. In 2023, a mutation in the THAP11 gene was linked to a new spinocerebellar ataxia[6].

The pathophysiological mechanisms depend on where in the gene the TNR expansion occurs and are basically divided into two: expansion in the coding region of the gene or expansion in the non-coding region[5]. When TNR expansion occurs in the coding region of the gene, the translated proteins are defective and form cytoplasmic protein aggregates. This mechanism is characteristic of Huntington's disease, which connects the disease with other neurodegenerative diseases such as Alzheimer's, Parkinson's, amyotrophic lateral sclerosis and spongiform encephalopathies[7]. In an experimental study, amyloid-like protein aggregates were formed, both in vivo and in vitro, due to proteolysis of the mutated protein encoded by exon 1 of the HD gene[8]. These protein aggregates are characteristic of PolyQ disorders. In some cases, it is possible to visualize these aggregates using light microscopy[8,9]. Most of the proteins that are translated from genes related to TNR expansion diseases, especially polyglutamine diseases, act as transcription factors in DNA[10]. Mutant TATA-Binding Protein (TBP) causes deregulation of gene transcription in two ways: because it is mutated, it loses affinity with common transcription factors that are needed to transcribe genes; at the same time, the

mutation increases affinity for factors that induce other transcriptions, sequestering them and interrupting their[9]. Although it is well established in the literature that the mutated protein is primarily responsible for the pathogenic mechanism in PolyQ disorders, studies suggest that the RNA transcript may also be related to the genesis of these neurodegenerative disorders[11]. Furthermore, it has been proven, although further studies are still needed, that the immune system plays a role in the pathogenesis of Huntington's disease. As for the other PolyQ disorders, studies are still needed to prove the involvement of the immune system[12]. In a way, the SCA1 disorder phenotype is directly linked to the pathophysiological processes of the aggregation of intracellular complexes, related to the ability of the polyglutamine tracts of the mutated protein to bind to other proteins such as PQBP1 and RNA pol II, causing transcriptional deregulation[10]. On the other hand, when expansion occurs in a non-coding region, instability or even silencing can occur in the gene in question[5]. In the case of FXS, this instability occurs in a gene located on one of the sex chromosomes: the X. Regarding the pathophysiological mechanisms of these molecular expansions, we can highlight at least two physiological consequences: gain-of-function and loss-of-function. While gain-of-function is related to increased protein translation and the formation of fibrillar aggregates, loss-of-function is related, for example, to gene silencing due to hypermethylation of the non-coding region[13]. The expanded PolyQ tracts synthesized by AUG-dependent canonical translation are believed to lead to neurodegeneration[13]. There is also a modality related to gain-of-function in expansions in the non-coding region: repeat-associated non-AUG translation. In these cases, there is protein translation independent of the AUG initiation codon[14].

Regarding the therapy of disorders caused by TNR expansions, the various peculiarities of specific disorders should be borne in mind. The β-sheet conformation of the PolyQ regions is responsible for cytotoxicity[13]. A peptide was discovered that prevents the transition of these proteins to the β-sheet conformation[13], consequently bringing therapeutic benefits. Also, as TNR expansions are recently discovered genetic disorders, treatment options are likely

to be quite limited. Early therapeutic intervention in cases of HD and SCA1 is related to a better prognosis[15]. Larger animal models may be more useful in analyzing the interactions between drugs and the body's selective barriers than the current mouse models, given the greater anatomical similarity between the larger species and the human species[16]. The notion that preventing protein aggregation in PolyQ diseases as a therapeutic mechanism has yet to be proven[17]. It is necessary to consider the various pathophysiological mechanisms for treatment.

The pathophysiology of disorders caused by TNR expansions have been hampered for many years due to the lack of efficient techniques for identifying and analyzing the affected genomic region. The development of simple techniques for analyzing DNA in the 1970s and 80s allowed the first disease related to TNR expansion to be described in the early 1990s, Fragile X syndrome[18]. The etiological definition of this disorder, whose main characteristic is intellectual disability, can be considered a major achievement, as it opened a field for new molecular investigations. In this way, TNR expansions were characterized as the basis for the molecular etiology of other disorders. TNR expansions have already been confirmed in the literature as being responsible for pathological alterations in the body. Over the 30 years since the relationship between TNR expansions and FXS, at least 48 other disorders have also been associated with TNR expansions in specific genes[5]. However, identifying TNR expansions in new genomic regions and determining the sizes of the expanded regions remains a challenge. The development of new approaches to DNA sequencing (NGS - Next Generation Sequencing) has the potential to revolutionize the diagnosis of various disorders, including those caused by TNR expansions

METHODS

We followed the recommendations of Page (2020) to carry out this integrative review. We searched PubMed (www.ncbi.nlm.nih.gov/pubmed) for scientific articles using the descriptors "trinucleotide expansion", "DNA analysis",

"genome" and "DNA sequencing". We listed the main diseases caused by TNR expansions and the possible mechanisms of the pathogenesis of these expansions and described how new DNA analysis techniques have enabled the description of new genomic regions with disease-related TNR expansions.

RESULTS AND DISCUSSION

We identified from the literature, 68 disorders related with nucleotide repeat disorders, 16 were excluded for having repeats of more than three nucleotides and five for not having clinical phenotypes. In total, 47 TNR expansion disorders have been selected (Table 1).

Table 1: Description of gene targets related to disorders by trinucleotide repeats (TNR) expansions.

Chrom. location	Gene	Gene name	Repeat unit	Disorder	Acronym
1q21	NOTCH2NLC	Notch 2 N-terminal like C	GGC	Essential tremor	ET
				Neuronal intranuclear inclusion disease	NIID
				Oculopharyngodistal myopathy 3	OPDM3
2q11	AFF3	ALF transcription elongation factor 3	CGG	Intellectual disability associated with fragile site FRA2A	FRA2A
2q31.1	HOXD13	Homeobox D13	GCG	Synpolydactyly type 1	SPD1
2q32.2	GLS	Glutaminase	GCA	Global developmental delay, progressive ataxia and elevated glutamine	GDPAG
3p21	ATXN7	Ataxin 7	CAG	Spinocerebellar ataxia – Type 7	SCA7
3q23	FOXL2	Forkhead box L2	GCN	Blepharophimosis, ptosis and epicanthus inversus syndrome	BPES
4p13	PHOX2B	Paired like homeobox 2B	GCN	Congenital central hypoventilation syndrome	CCHS
4p16.3	HTT	huntingtin	CAG	Huntigton disease	HD
5q3	PPP2R2B	Protein phosphatase 2 regulatory subunit beta	CAG	Spinocerebellar ataxia type 12	SCA12

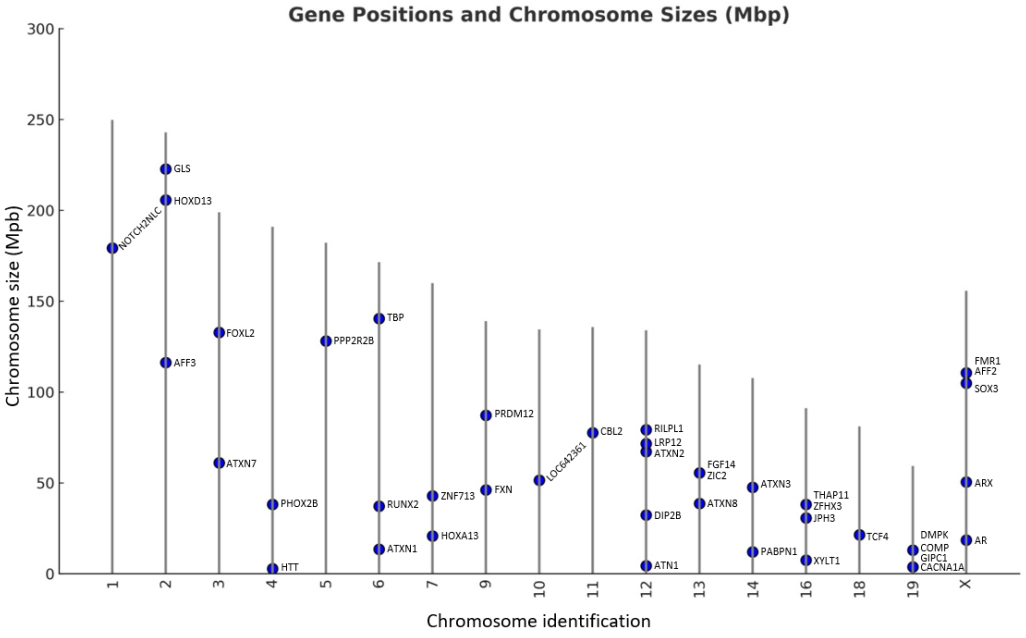
6p21	RUNX2	RUNX family transcription factor 2	GCN	Brachydactyly and cleidocranial dysplasia	BCCD
6p22	ATXN1	Ataxin 1	CAG	Spinocerebellar ataxia – Type 1	SCA1
6q27	TBP	TATA-box binding protein	CAG	Spinocerebellar ataxia – tipo 17	SCA17
7p11.2	ZNF713	Zinc finger protein 713	CGG	Autism spectrum disorder associated with fragile site fra7a	ASD/FRA7A
7p15.2	HOXA13	Homeobox A13	GCN	Hand-foot-genital syndrome	HFGS
9q21.11	FXN	Frataxin	GAA	Friedreich ataxia	FRDA
9q34.12	PRDM12	PR/SET domain 12	GCN	Hereditary sensory and autonomic neuropathy type viii	HSAN VIII
10q22.3	LOC642361 / NUTM2B-AS1	NUTM2B antisense RNA 1	CGG	Oculopharyngeal myopathy with leukoencephalopathy	OPML 1
11 - Deletion of long arm	CBL2	Cbl proto-oncogene 2	CCG	Jacobsen syndrome	JBS
12p13.31	ATN1	Atrophin 1	CAG	Dentatorubro pallidoluysian atrophy	DRPLA
12q13.1	DIP2B	Disco interacting protein 2 homolog B	CGG	Intellectual disability associated with fragile site FRA12A	FRA12A
12q24	ATXN2	Ataxin 2	CAG	Spinocerebellar ataxia – Type 2	SCA2
12q24.3	LRP12	LDL receptor related protein 12	CGG	Oculopharyngodistal myopathy 1	OPDM 1
12q24.31	RILPL1	Rab interacting lysosomal protein like 1	CGG	Oculupharyngodistal myopathy 4	OPDM 4
13q21	ATXN8	Ataxin 8	CAG	Spinocerebellar ataxia – Type 8	SCA8
13q32	ZIC2	Zic family member 2	GCN	Holoprosencephaly type 5	HPE5
13q33.1	FGF14	Fibroblast growth factor 14	GAA	Spinocerebellar ataxia – tipo 50	SCA50/ ATX-FGF14
14q11.2	PABPN1	Poly(A) binding protein nuclear 1	GCG	Oculopharyngeal muscular dystrophy	OPMD
14q32.1	ATXN3	Ataxin 3	CAG	Spinocerebellar ataxia – Type 3	SCA3
16p12.3	XYLT1	xylosyltransferase 1	CGG	Baratela–scott syndrome	BSS
16q22	ZFHX3	Zinc finger homeobox 3	GGC	Spinocerebellar ataxia 4	SCA4
16q22.1	THAP11	THAP domain containing 11	CAG	Spinocerebellar ataxia novel subtype	SCA

16q24.2	JPH3	Junctophilin 3	CAG	Huntington disease-like 2	HDL2
18q21.2	TCF4	Transcription factor 4	CTG	Fuchs endothelial corneal dystrophy	FECD
19p13	CACNA1A	Calcium voltage-gated channel subunit alpha1 A	CAG	Spinocerebellar ataxia – Type 6	SCA6
19P13.1	COMP	Cartilage oligomeric matrix protein	GAC	Pseudoachondroplasia and multiple epiphyseal dysplasia	PSACHMED
19p13.12	GIPC1	GIPC PDZ domain containing family member 1	CGG	Oculopharyngodistal myopathy 2	OPDM2
19q13.32	DMPK	DM1 protein kinase	CTG	Myotonic dystrophy type 1	DM1
Xp21.3	ARX	Aristaless related homeobox	GCN	Early infantile epileptic encephalopathy type 1	EIEE1
				X-Linked intelectual disability	XLID
Xq12	AR	Androgen receptor	CAG	Spinal and Bulbar Muscular Atrophy	SBMA
Xq26.3	SOX3	SRY-box transcription factor 3	GCN	Mental retardation with isolated growth hormone deficiency	MRGH
Xq27.3	FMR1	Fragile X messenger ribonucleoprotein 1	CGG	Fragile X syndrome	FXS
				Fragile X-associated premature ovarian infertility/insufficiency	FXPOI
				Fragile X-associated tremor/Ataxia syndrome	FXTAS
Xq28	AFF2	ALF transcription elongation factor 2	CCG	Fragile X syndrome	FRAXE

Research into the molecular etiology of these disorders revealed that they are caused by trinucleotide expansions in 42 different genomic targets. In 39 out of the 42 genomic targets, TNR expansions in a determined gene are associated with a single disorder. However, in 3 out of the 42 genomic targets, TNR expansions in different locations of each gene are related to more than one disorder (Table 1; ARX, FMR1 and NOTCH2NLC genes). Thus, mutations in different regions of the ARX gene cause the disorders "X-Linked Intellectual Disability" and "Early Infantile Epileptic Encephalopathy Type 1"; mutations in the NOTCH2NLC gene cause "Neuronal Intranuclear Inclusion Disease", "Oculopharyngodistal Myopathy 3" and "Essential Tremor"; and mutations in the FMR1 gene cause "Fragile X Syndrome", "Fragile X-Associated Premature Ovarian Infertility/Insufficiency" and "Fragile X-

Associated Tremor/Ataxia Syndrome". Of all 42 genomic targets, 37 are located on autosomal chromosomes and 5 are located on sex chromosomes, in this case the X chromosome (Figure 1).

Figure 1: Distribution and position of targets associated with TNR expanssions.



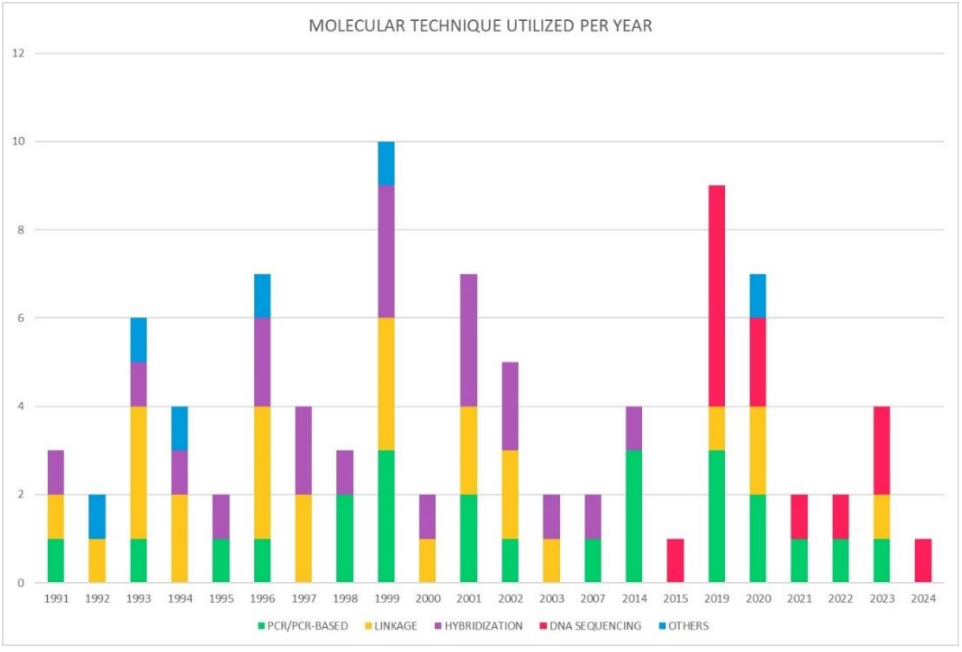
Legend: Only chromosomes with a described TNR target are shown. Complete gene names are presented in table 1. Chromosome sizes presented in Mb (mega base pairs).

Estimated target density in human genome is 1.3 targets/100 Mb. Chromosome 19 had the highest target density of all those analyzed, with 6,83 targets/100 Mb (Figure 1). Structurally, chromosome 19 has several cytosine and guanine islands. This structure is conserved in primates that are separated by 70 million years of evolution. At the same time as this structure is highly conserved, these islands predispose to various polymorphisms[19]. In addition, chromosome 19 has other unique characteristics, such as twice the gene density of the average of the other chromosomes[19].

The main techniques used to identify genomic targets in which TNR expansions are related to human disorders were Linkage – 25 targets; PCR and

PCR-based techniques - 24 targets; Hybridization techniques (CGA, ExSc, FISH) - 21 targets; DNA sequencing (LRS, WGS, Sanger) - 14 targets; Other techniques (Karyotyping, cloning) – 6 targets (Figure 2).

Figure 2: Techniques used to identify trinucleotide repeat disorders over time.



More than one technique can be used to identify a single trinucleotide repeat disorder. The molecular techniques are classified as PCR/PCR-BASED in green, Linkage in yellow, Hybridization in purple, DNA sequencing in red, and Others in blue.

Sequencing techniques can be divided into first generation (Sanger), second generation (Next Generation Sequencing) and third generation (long-read sequencing) techniques. In terms of the type of DNA sequencing technique used to identify disorders caused by TNR expansions, we have first generation (Sanger; 2 targets); second generation (Exome Screening (5 targets) and WGS (4 targets); and third generation (LRS 4 targets). Between 2005 and 2018, there was a stagnation in the description of new targets, which resumed in 2019, mainly due to next-generation sequencing (NGS) techniques. The impact of the molecular techniques on the discovery of TNR expansion disorders are as follows: NGS – 14%;

Hybridization – 24%; Linkage – 28%; PCR/PCR-BASED – 27%; Others – 7%. In the 33 years since the discovery of TNR expansion diseases, during the last 9 years, we have had 13 disorders in which DNA sequencing was fundamental for the identification of the target affected. During the first 24 years, we had 38 disorders identified by all other techniques combined (Linkage, PCR/PCR-Based, Hybridization, Others). Proportionally, we identified more disorders in the last 9 years (9/13 = 0.69) with the DNA Sequencing Technique than in the first 24 years (24/38 = 0.63) with all the molecular techniques combined. This is a major indication of how NGS is contributing to the discovery of new targets related to TNR expansion disorders and the importance of this technique.

All these findings are important from a clinical point of view, especially in relation to early diagnosis and the development of therapies for disorders associated with TNR expansions. As a methodological limitation, there was no peer review of the articles and only one author was responsible for screening and selecting the articles.

ACKNOWLEDGMENTS

We would like to thank PIBIC/FAPEMIG for funding this research.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest that could be perceived as prejudicial to the impartiality of the reported research.

AUTHORS CONTRIBUTIONS

A.L.P conceived the study. L.R.B collected the data. A.L.P and L.R.B. performed data analysis and wrote the manuscript.

REFERENCES

1. Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze AV, Mikheenko A, Vollger MR, Altemose N, Uralsky L, Gershman A, et al. The complete sequence of a human genome. Science. 2022; 376(6588):44-53. https://doi.org/10.1126/science.abj6987

2. Liehr T. Repetitive Elements in Humans. Int J Mol Sci. 2021; 22(4):2072. https://doi.org/10.3390/ijms22042072

3. Carpenter NJ. Genetic anticipation. Expanding tandem repeats. Neurol Clin. 1994; 12(4):683-97. https://doi.org/10.1016/S0733-8619(18)30071-9

4. Yau WY, O'Connor E, Sullivan R, Akijian L, Wood NW. DNA repair in trinucleotide repeat ataxias. FEBS J. 2018; 285(19):3669-3682. https://doi.org/10.1111/febs.14644

5. Depienne C, Mandel JL. 30 years of repeat expansion disorders: What have we learned and what are the remaining challenges? Am J Hum Genet. 2021; 108(5):764-785. https://doi.org/10.1016/j.ajhg.2021.03.011

6. Tan D, Wei C, Chen Z, Huang Y, Deng J, Li J, Liu Y, Bao X, Xu J, Hu Z, et al. CAG Repeat Expansion in THAP11 Is Associated with a Novel Spinocerebellar Ataxia. Mov Disord. 2023; 38(7):1282-1293. https://doi.org/10.1002/mds.29412

7. Bates GP, Dorsey R, Gusella JF, Hayden MR, Kay C, Leavitt BR, Nance M, Ross CA, Scahill RI, Wetzel R, Wild EJ, Tabrizi SJ. Huntington disease. Nat Rev Dis Primers. 2015; 1:15005. https://doi.org/10.1038/nrdp.2015.5

8. Scherzinger E, Lurz R, Turmaine M, Mangiarini L, Hollenbach B, Hasenbank R, Bates GP, Davies SW, Lehrach H, Wanker EE. Huntingtin-encoded polyglutamine expansions form amyloid-like protein aggregates in vitro and in vivo. Cell. 1997; 90(3):549-58. https://doi.org/10.1016/s0092-8674(00)80514-0

9. Liu Q, Pan Y, Li XJ, Li S. Molecular Mechanisms and Therapeutics for SCA17. Neurotherapeutics. 2019; 16(4):1097-1105. https://doi.org/10.1007/s13311-019-00762-z

10. Figueiredo AS, Loureiro JR, Macedo-Ribeiro S, Silveira I. Advances in Nucleotide Repeat Expansion Diseases: Transcription Gets in Phase. Cells. 2023; 12(6):826. https://doi.org/10.3390/cells12060826

11. Ravache M, Abou-Sleymane G, Trottier Y. Les maladies neurodégénératives par expansion de polyglutamine: physiopathologie et stratégies thérapeutiques [Neurodegenerative polyglutamine expansion diseases: physiopathology and therapeutic strategies]. Pathol Biol. 2010; 58(5):357-66. French. https://doi.org/10.1016/j.patbio.2009.12.004

12. Olejniczak M, Urbanek MO, Krzyzosiak WJ. The role of the immune system in triplet repeat expansion diseases. Mediators Inflamm. 2015; 2015:873860. https://doi.org/10.1155/2015/873860

13. Fujino Y, Nagai Y. The molecular pathogenesis of repeat expansion diseases. Biochem Soc Trans. 2022; 50(1):119-134. https://doi.org/10.1042/BST20200143

14. Zhou ZD, Jankovic J, Ashizawa T, Tan EK. Neurodegenerative diseases associated with non-coding CGG tandem repeat expansions. Nat Rev Neurol. 2022; 18(3):145-157. https://doi.org/10.1038/s41582-021-00612-7

15. Thompson LM, Orr HT. HD and SCA1: Tales from two 30-year journeys since gene discovery. Neuron. 2023; 111(22):3517-3530. https://doi.org/10.1016/j.neuron.2023.09.036

16. Lee LKC, Leong LI, Liu Y, Luo M, Chan HYE, Choi CHJ. Preclinical Nanomedicines for Polyglutamine-Based Neurodegenerative Diseases. Mol Pharm. 2021; 18(2):610-626. https://doi.org/10.1021/acs.molpharmaceut.0c00506

17. Jones L, Houlden H, Tabrizi SJ. DNA repair in the trinucleotide repeat disorders. Lancet Neurol. 2017; 16(1):88-96. https://doi.org/10.1016/S1474-4422(16)30350-7

18. Verkerk AJ, Pieretti M, Sutcliffe JS, Fu YH, Kuhl DP, Pizzuti A, Reiner O, Richards S, Victoria MF, Zhang FP, et al. Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell. 1991; 65(5):905-14. https://doi.org/10.1016/0092-8674(91)90397-h

19. Harris RA, Raveendran M, Worley KC, Rogers J. Unusual sequence characteristics of human chromosome 19 are conserved across 11 nonhuman primates. BMC Evol Biol. 2020; 20(1):33. https://doi.org/10.1186/s12862-020-1595-9