RESEARCH ARTICLE

Trinucleotide repeat diseases - antecipation diseases

Inês Lopes Cardoso, Vicente Marques

Cardoso IL, Marques V. Trinucleotide repeat diseases - antecipation diseases. J clin gen genomics. March-2018;1(1):4-9.

Dynamic mutations involve expansion of the number of repeat units consisting of three or more nucleotides in tandem (i.e. adjacent to one another) present in a gene or in its neighborhood. These repeats may occur in different genes and may code for different aminoacids.
According to expansions sizes, it is possible to have unaffected individuals that are carriers of a pre-mutation. Instability of triplet repeat size can lead to gradual expansion through generations, a phenomenon called anticipation. Genetic anticipation is characterized by the reduction in the age of disease onset and by a worsening of symptoms in affected individuals in successive generations.
This work describes dynamic mutations giving emphasis on triplet repeats diseases, making the parallel with disease anticipation. Treatment strategies that have been developed during the last years are also discussed.

KEYWORDS: Trinucleotide repeats; Anticipation diseases.

INTRODUCTION

Eukaryotic genomes contain several types of repetitive sequences like long repeats, satellite DNA and many other sequences of diverse sizes and repetitive levels not yet classified [1,2]. Repeats of simple sequences or microsatellites are present in high abundance in many prokaryotic and eukaryotic genomes [3,4]. According with McMurray [5] repetitive DNA sequences constitute 30% of the human genome and, in most species, changes in the length of repetitive DNA during evolution give rise to species diversity.

In the beginning of 90th decade, the identification of a novel class of mutations leading to diseases gave rise to considerable enthusiasm in the scientific community [6]. This type of mutations is the presence of trinucleotide repeats that belong to the family of microsatellites also known as repeat sequences in tandem. This type of repetitive DNA shows high degree of mutation, being present in prokaryotic and eukaryotic genomes [2,7].

Genetic instability of triplet repeats is a trait present in certain types of hereditary neurological diseases [8]. There are around twenty neurologic and neuromuscular diseases caused by trinucleotides repeat expansions, but their development mechanisms are still not known [5,9]. Triplet repeat disorders include Myotonic Dystrophy types (1, 2, 3, 6, 7 and 8), Spinocerebellar Ataxia, Huntington Disease, X Fragile Syndrome, Friedreich Ataxia, Haw River Syndrome and Jacobsen Syndrome. The cause of these disorders has been described as the expansion of the number of triplet repeats CTG-CAG, CGG-CCG or GAA-TTC in associated genes [9,10]. Symptoms will be present as soon as the individual accumulates a determined number of repeats that results from the somatic instability of the expanded triplet. Long et al. [11] determined the extent of somatic instability of the GAA repeats in several tissues from 15 patients with Friedreich's ataxia. In this work, researchers observed significantly longer GAA tracts detected in heart and pancreas. Moreover, analysis of the GAA repeat length in lymphocytes collected over a span of 7-9 years, showed progressive expansions of that repeat with maximum gain of approximately 9 repeats per year [11].

This instability of the expanded triplet is responsible for the increase in the number of repeats from one generation to the next. Consequently, the age of symptoms onset in these disorders follows an unusual genetic pattern called anticipation, where the disease becomes more severe and assumes earlier expression in each generation [9,10]. This anticipation phenomenon has been described in genetically complex disorders, like bipolar affective disorder, schizophrenia, Crohn's disease and psoriasis [12,13]. All triplet repeat disorders show anticipation [14] and a significant correlation between age at onset of the disease and length of the expanded repeat is observed [11,15].

Despite last decade efforts of researchers in the development of treatments for unstable repeat expansions, triplet repeat diseases are still not curable [6].

MATERIALS AND METHODS

The present revision was done through search of scientific publications, revision papers, books and thesis. The search was performed using the websites: Pubmed, B-On, SciELO, Science Direct, as well as libraries of O´Porto Medicine Faculty (FMUP) and Fernando Pessoa University.

Keywords used in the search were: "human genome", "genetic anticipation", "diseases of trinucleotides repeats", "neurodegenerative disorders", "spinocerebellar ataxia", "Huntington disease", "X fragile disease", "myotonic dystrophy", "Friedrich ataxia", "X fragile associated ataxia".

DYNAMIC MUTATIONS

Mutations are permanent cell genome changes in a certain DNA sequence that are transmitted to daughter cells. These changes can be of a single nucleotide or more extensive.

Dynamic mutations involve the expansion of the repetitive unit number present in a certain gene or in its neighborhood, consisting of three or more nucleotides in tandem. The presence of these repetitive sequences makes more likely the occurrence of DNA polymerase slippage, leading to the insertion of extra nucleotides in that area during DNA replication [13].

Tandem repeats are usually classified in microsatellites (1-6 bps), minisatellites (6-24 bps, and in certain cases longer) and classic satellites. These repeats have been found in regulatory regions of eukaryotic genes, interacting with transcription factors [4].

The extent of triplet repeats propagation varies from genome to genome, and results from mistakes and rearrangements during replication. In this perspective, a big portion of repetitive genomic DNA can be considered as having no role, however the presence of tandems can be related with a biological function. As an example, long tandem repeats represent a big

Faculdade de Ciências da Saúde, Universidade Fernando Pessoa, Rua Carlos da Maia, 296, 4200-150 Porto, Portugal

**Correspondence: Inês Lopes Cardoso, Faculdade de Ciências da Saúde, Universidade Fernando Pessoa, Rua Carlos da Maia, 296, 4200-150 Porto, Portugal , Telephone: 351 225074630/ 965772845; E-mail: mic@ufp.edu.pt*

Received: January 22, 2018, Accepted: February 25, 2018, Published: February 28, 2018

function in heterochromatin being involved in the formation and function of the centromere [1].

Polymorphic triplet repeats are better tolerated than repeats of untranslated sequences of dinucleotides and tetranucleotides because the change does not affect sequence open reading frame [7].

In normal individuals, chromosomes have a certain number of triplet repeats (usually 5-37), that are polymorphic in the general population [10]. The increase in the number of triplet repeats to 50-120 is associated to some diseases/syndromes [8].

According to Krzyzosiak et al. [6] each disease is associated with mutations in only one gene that would trigger a pathogenic process through aberrant gene expression and toxic properties of mutant proteins. The same work revealed that mutant proteins and corresponding RNAs can act in parallel exhibiting the pathogenic effects in an independent fashion [6].

TRINUCLEOTIDES REPEATS DISEASES

Triplet repeat diseases constitute the most diverse and interesting class of microsatellites present in genomic sequences [16,17]. From the 64 triplets of nucleotides, 4 are classified as mononucleotides repeats (AAA, TTT, GGG, or CCC) [16]. Triplet repeat AGC is considered similar to GCA or CAG. The same happens with all other triplet repeats [16].

Density and distribution of triplet repeats

Subramanian and collaborators [3] analyzed human DNA sequences using a program designed to detect the abundance and distribution of triplet repeats in coding and non-coding portions of the genome. Repeat density was calculated for each type of triplet (bp/Mb) in each chromosome. Chromosome 1 showed the biggest number of triplet repeats while chromosome Y showed the lowest. Repeat AAT was the one having the highest density in the whole genome. The size average of this repeat was 282 bp/Mb per sequence. In second place were AAC (average of 193 bp/Mb), AAG and AGG (both with 77 bp/Mb). ACG repeat registered the lowest density (average varied between 0 and 1 bp/Mb) in most chromosomes, followed by ACT as the second less dense with an average of 9 bp/Mb [3].

The number of repeats is inversely proportional to occurrence. It was observed the following distribution of triplets repeats: AAT (26 037), AAC (18 707), followed by AGG (10 245) that registered highest number of occurrences for 4 repeats in tandem [3].

Distribution of triplet repeats in several genomic regions

Analysis of triplet repeats type occurrence was conclusive in terms of abundance of repeats in introns compared with the occurrence in exons or untranslated regions [3,16].

AAT and AAC repeats were the most abundant and ACG and ACT repeats revealed poor abundance in all analyzed regions [3,16]. In exons, triplet repeats, like AGC, AGG and CCG, are respectively the most abundant, while in untranslated regions, triplets like CCG and AAT followed by AAC showed predominant occurrence. However, these levels are very low when compared with its levels in exonic regions [3,16].

Association of triplet repeats with genes

Subramanian et al. [3] showed that around 2135 genes contained at least one type of triplet in exons having a minimum of four repeats in tandem. From these 2135 genes, 171 had length repeats of at least 30 bps. The same study revealed that genes had mainly AGC and CCG repeats. Moreover, repeats AAC, AAT, AAG, ACC and ATC are not predominantly found in genes, and ACT or ACG repeats are completely absent in genes [3,16].

Triplet repeats and the promoter region

Analysis of 500 bp gene promoter sequences revealed the existence of preferred association of certain repeats with these regulatory sequences. Most promoter regions contain CCG tandem repeats [3]. The second most abundant type of repeat in promoter regions is AGG followed by AAT and AAC [3,16]. The abundance of the last two repeats can be explained on the basis of its abundance in the whole genome, but this is not true for CCG and AGG repeats [3]. For this reason, it is not easy to explain the predominance of CCG triplets in promoters, however as promoter regions possess CpG islands, it is expected that most promoter sequences have CCG repeats [3].

From the 269 genes that had at least one triplet repeat in the promoter region, none had ACT repeats [3].

Triplet repeats in human transcription

The big abundance of some triplet repeats in cells genome raises questions about the roles of these sequences in transcription. Triplet repeats differ in length, and the level of gene expression depends on the secondary structures that these regions can form as well as on the presence of repeats that disrupt the homogeneity of the repeat [6,18].

In fact, a study showed that the human genome includes more than 32000 extensions of uninterrupted triplet repeats composed of six or more repeat units [7]. Even human exons, that represent less than 3% of genomic sequences, show a maximum of 1030 extensions of triplet repeats. These were identified in exons of 878 genes [6]. Triplet repeats that are highly present in exons are CNG (where N represents any nucleotide), CGA and AGG [6,18]. CTT, CAT, CAA, TAA and TTA repeats are present in lower amounts [6,18].

Triplet repeat sequences rich in AT, like CTT, AAC and AAT are particularly sub-represented in exons while repeats rich in GC (CGG, CAG and CCG) are highly abundant suggesting a functional role of these sequences [7,16]. Moreover, shorter sequences are more prevalent, and the triplet repeat frequency decreases exponentially with its length [6].

In exons, 60% of triplet repeat sequences are located in the open reading frame and repeat codons are translated into poly-glutamine, poly-alanine, poly-glutamate and poly-leucine [7]. CCA, CAG, CTG, CCT, AGG, AAG and ATG triplets appear frequently in the open reading frame region [6]. Besides these, 28% of gene repeats are located in the 5' untranslated region (CCG and CGG are the most frequent with an abundance of 52% and 62%, respectively) and 13% in the 3' untranslated region (the most abundant are triplet repeats rich in AT) [6].

Functional studies have been done to clarify the roles of triplet repeats present in genes. Krzyzosiak and collaborators [6] saw that these genes code for proteins involved in transcription or proteins that interact with nucleic acids.

According to Sobczak et al. [7] triplet repeats can play an important role in the regulation of several cellular processes, like DNA repair during replication and diverse stages in gene expression.

RNA structural types in triplet repeats

Functional analysis of triplet repeats in RNAs, through oligoribonucleotides structures (ORNs) formed by several specific triplet repeats has been investigated in different experimental conditions. The first types of ORNs analyzed was GNC. All these ORNs assumed stable hairpin structures [6,18]. The hairpin formed with GNC repeats was composed of periodic repetition of base pairs C-G and G-C, and the loop of the hairpin was formed by any four or seven nucleotides [6].

Except for CGG hairpin, all other CNG repeats form slippery hairpins, meaning that they tend to form alternative alignments unless a conformation by G-C clamp is present [6].

A comparative analysis of 20 triplet repeats showed that AGG and UGG tend to form quadruplex structures that are considered the most stable ones, followed by CGA, CGU and CNG repeats that give rise to hairpins

that are more stable than the ones formed by UAG, AUG, UUA, CUA and CAU repeats. CAA, UUG, AAG, CUU, CCU, CAA and UAA repeats do not form any secondary structure [7,18]. The analysis of hairpin and quadruplex repeats, with the help of biophysical methods revealed the following order of CNG repeat stability: CGG>CAG>CUG>CCG. The stability of quadruplex structures of AGG and UGG triplets is similar [7,18].

Recent work studied the structure and dynamics of RNA repeat expansions responsible for the development of Huntington disease and Myotonic Dystrophy type 1 [19]. Researchers analysed the structures of two RNA constructs containing three copies of r(CAG) or r(CUG) motif by nuclear magnetic resonance spectroscopy. The 1×1 internal loops of r(3×CAG) were stabilized by one-hydrogen bond AA pairs, while those of r(3×CUG) preferred one or two hydrogen bond UU pairs [19]. Chen et al. [19] observed that the loop regions of RNA structures are dynamic.

Triplet repeat and its interaction with proteins

Research on proteins resulting from trinucleotide sequences present in triplet repeats disorders has a lot of interest in the study of toxic features of mutant transcripts [20]. Performed studies were mainly focused on the identification and structural characterization of proteins resulting from triplet repeats [20].

Cellular proteins that are involved in the formation or processing of hairpins are pointed out as one of the causes of triplet repeats expansion [20].

Using mouse models, it was seen that the MutSβ protein, responsible for identifying mismatches, is the causing agent of triplet expansion in Huntington disease and Myotonic Dystrophy 1 [20,21]. Most inheritable somatic expansions are blocked by knockout of Msh2 or Msh3 genes that code for the two subunits of MutSβ protein [20,21]. Studies performed in mice, showed that mismatch repair done by complex Msh2-Msh3 is essential for CAG repeat expansion in germ line and somatic cells [22]. This supports the hypothesis of mismatch repair being an important source of expansion especially in non-proliferative tissues where replication is low [20,21].

Types of triplet repeat disorders

The number of neurodegenerative disorders resulting from triplet repeats is increasing, as well as the complexity of its molecular pathogenesis [23].

More than 20 different genes containing unstable triplet repeats are responsible for the pathogenesis of human neurological diseases. Repeats like CTG, CGG, GAA and CAG, are source of degenerative changes leading to Myotonic Dystrophy type 1 (MD1), Fragile X associated ataxia (FXTAS), Friedreich ataxia (FRDA), as well as Huntington disease (HD) and several types of spinocerebellar ataxias (SCAs) [6,24].

There are two main types of triplet repeats diseases. The first type corresponds to the group of disorders where the expanded triplet repeat is translated, usually including poly-glutamine or poly-alanine homorepeats.

In the second type, triplet repeats are present in non-coding regions of the gene. These repeats can be expressed at the mRNA level (affecting expression levels of coded proteins) but are not translated, having no effect on protein structure. However, sometimes, the produced mRNA can be translated, giving rise to highly repetitive peptides of different length and aminoacid content. This might be the crucial step to the formation of a toxic homopolypeptide that can lead to the development of the pathology.

Type 1: Diseases of translated triplet repeats

Type 1 of triplet repeats diseases involves the translation of the expanded sequence leading to the synthesis of a chain of glutamines or alanines. This results in the formation of protein aggregates inside the cell [6,23]. These aggregates lead to the appearance of the toxic function of the mutant protein [23].

In the case of poly-glutamine repeats, the CAG repeat is highly unstable and consequently there is a high level of polymorphism among affected individuals and in different tissues.

These proteins lead to highly stable β-rich amyloid-like protein inclusions [25], that can accumulate in the nucleus or the cytoplasm. Since these poly-glutamine expansion proteins usually interact with several other proteins, the produced aggregates contain diverse proteins like chaperones, ubiquitin and other poly-glutamine containing proteins, making them unable to perform their functions [25]. As soon as they form large inclusions, these proteins become resistant to degradation. These expansion proteins are usually involved in the regulation of neurogenesis or transcription [26].

Concerning poly-alanine repeats, the extension of these tracts occurs by unequal homologous recombination during meiosis, while other disease expansions extend because of DNA polymerase slippage [27]. These extensions result in structural changes in the repeat peptide that is dependent on its length. Big repeat length usually pushes the peptide to degradation or aggregation. Longer repeats tend to aggregate due to the formation of stable β-sheets, that are resistant to degradation [28]. These dense aggregates mis-localize to the cytoplasm and can contain the mutant as well as the wildtype peptide, precluding its normal function.

Most proteins having pathogenic poly-alanine repeat expansions act as transcription factors, except for PABP-2 that controls the length of polyadenylate tails added during mRNA processing.

The number of triplet repeats that leads to the development of the pathology differs from gene to gene. Moreover, in most cases the repeated sequence shows somatic instability, meaning that it has a strong probability of expansion. The speed of expansion of these triplet repeats also varies according with the gene.

Table 1 lists a group of this type of disorders with the corresponding repeat sequence, its location and the length of the repeat that leads to disease development.

Table 1 Localization, triplet repeat sequence, its location and the length of the repeat that leads to type 1 repeat diseases.

Disease (poly-glutamine)	Sequence	Number of repeats (normal)	Number of repeats (disease)	Localization
Dentatorubral-pallidoluysian atrophy (DRPLA)	CAG	6-35	49-88	ATN1 (exon 5)
HD	CAG	6-29	38-180	HTT (exon 1)
SCA1	CAG	6-39	41-83	ATXN1 (exon 8)
SCA2	CAG	<31	32-200	ATXN2 (exon 1)
Machado-Joseph disease (SCA3)	CAG	12-40	52-86	ATXN3 (exon 8)
SCA6	CAG	<18	20-33	CACNA1A (exon 47)

SCA7	CAG	4-17	>36 to >460	ATXN7 (exon 3)
SCA17	CAG	25-42	45-66	TBP (exon 3)
Spinal Muscular Bulbar Atrophy (SMBA)	CAG	13-31	40	AR (exon 1)
Disease (poly-alanine)	Sequence	Number of repeats (normal)	Number of repeats (disease)	Localization
Synpolydactyly (SPD II)	GCG	15	>21	HOXD13 (exon 1)
Hand-foot genital syndrome (HFGS)	GCG	12	>17	HOXA13 (exon 1)
Cleidocranial dysplasia (CCD)	GCG	17	>26	RUNX2 (exon 1)
Holoprosencephaly disorder (HPE)	GCG	9	-	ZIC2 (exon)
Congenital central hypoventilation syndrome (CCHS)	GCG	20	-	PHOX2B (exon 3)
ARX-nonsyndromic X-linked mental retardation (XLMR)	GCG	16	>17	ARX (exon 2)
Oculopharyngeal Muscular Dystrophy (OPMD)	GCG	10	>11	PABPN1 (exon 1)

Type 2: Diseases of untranslated triplet repeats

In type 2 repeats disorders the triplet sequence is present in non-coding regions of genes [23] and can be expressed at mRNA level. These mRNAs form stable structures that can damage cells. But their main toxicity comes from the fact that although they are not translated, can affect the expression levels of coded proteins, since mutant RNAs can bind specific proteins and inhibit their normal functions. This model has been proposed for several triplet repeat diseases such as Myotonic Dystrophy type 1 [29] and Fragile X Tremor Ataxia Syndrome [30,31].

Another proposed disease mechanism is supported by the fact that these expansion repeats, although start codons are absent, sometimes suffer aberrant translation (repeat-associated non-AUG (RAN) translation), giving rise to toxic polypeptides with highly repeated sequences [30-32]. These peptide products may contribute to cell death through a gain of function toxicity.

Table 2 summarizes the expanded sequence, location and the average number of repeats required for the development of several untranslated triplet repeat disorders.

Table 2 Localization, repeat sequence and number of repeats that leads to several untranslated triplet repeats diseases.

Disease	Sequence	Number of repeats (normal)	Number of repeats (disease)	Localization
Myotonic Dystrophy type 1 (MD1)	CTG	5-37	<50	DMPK (3′ UTR)
MD2	CCTG	<30	75-11000	CNBP (intron 1)
Fragile X mental retardation (FRAX-E)	GCC	4-39	>200	AFF2 (5′ UTR)
Friedreich ataxia (FRDA)	GAA	5-30	70-1000	FXN (intron 1)
Fragile X Syndrome (FXS)	CGG	6-50	200-4000	FMR1 (5′ UTR)
Huntington disease-like 2 (HDL2)	CTG	6-27	36-57	JPH3 (exon 2A)
Spinocerebellar ataxia type 8 (SCA8)	CTG	15-34	89-250	ATXN8OS (3′ UTR)
SCA10	ATTCT	10-29	400-4500	ATXN10 (intron 9)
SCA12	CAG	7-28	66-78	PPP2R2B (5′ UTR)

Experimental Therapies directed against Triplet Repeats

Development of selective therapeutic agents is based on differences between normal and mutant transcripts concerning repeat sequence length and assumed secondary structures [6].

The most promising approaches of possible treatments aim lowering mutant protein by targeting DNA or RNA transcripts.

RNA targeting has been developed through three main strategies directed against triplet RNA repeats: RNA interference (RNAi), antisense oligonucleotides (ASOs) or using splicing inhibitors.

The first one is based on mutant transcripts degradation using interfering RNA (iRNA) since these RNA molecules bind to mRNA in the cytoplasm, leading to its degradation by the RNAse enzyme argonaute 2 [33]. This strategy has been applied in treatment of Huntington disease, and it was seen that a single treatment might permanently lower the levels of huntingtin [34].

ASOs act in a similar way, catalysing degradation of mRNA transcripts by RNAse H, and consequently leading to reduction of coded protein [6]. Kordasiewicz et al. [35] observed an 80% reduction in huntingtin transcript levels using this strategy [35].

According to the mechanism of action, these antisense reagents are called Cutters since they attach to complementary targets and induce their cleavage.

Splicing inhibitors have been used in the treatment of spinal muscular atrophy. The purpose of these molecules is to increase the synthesis of SMN protein by stopping alternative splicing that leads to truncated SMN protein and the development of neuromuscular degeneration [36].

Targeting DNA with the goal of reduction of protein expression can be performed through two approaches: zinc finger proteins and CRISPR/Cas9.

Zinc finger proteins form a structural motif able to bind DNA and hence might reduce coded protein levels. This strategy has been used to reduce levels of mutant huntingtin protein in animal models [37]. However, since these proteins lead to the formation of non-native molecules, there is the chance of development of immune reactions [37].

CRISPR/Cas9 is present in the bacterial immune system and its role is to cleave foreign DNA. In a recent study, this novel strategy was used and allowed to recover normal phenotype in myogenic cells derived from fibroblasts of patients with Myotonic Dystrophy type 1 [38]. The aim of the developed strategy was to eliminate the toxic mutant repeats. Researchers used the CRISPR/Cas9 gene-editing system, and were able to remove the repeat expansions, therefore preventing nuclear foci formation and splicing alterations [38]. The advantage of this approach when compared with previously described strategies is that affected cells can be permanently reverted to a normal phenotype. Permanent inactivation of Huntington's disease mutation was also achieved in fibroblasts of a patient using this approach [39].

In some triplet repeat diseases, the expanded repeat leads to reduction of gene expression. This is observed in patients having Friedreich's ataxia or spinal muscular dystrophy, that show a reduction in the expression of respectively FXN or SMN proteins. ASOs have also been used in cell lines of patients having Friedreich's ataxia and spinal muscular dystrophy to increase gene expression [40,41]. The proposed mechanism of action of ASOs in increasing gene expression, involves the binding of these compounds to the expanded repeat and expected blockage of the expanded RNA that inhibit recognition of DNA, prevent R-loop formation and allow more gene expression [40].

An important issue in therapy directed to triplet repeats is the need for gene and allele selectivity. Although there are numerous mRNA containing triplet repeats in the human transcriptome, the specific inhibition of mutant gene expression having as target certain repeat regions is a promising therapy strategy for these kind of diseases [6]. Agents have been developed that are able to distinguish between RNA hairpin structures formed by wildtype and mutant RNAs, since they differ in length and consequently diverge in stability. Hu et al. [42] used peptide nucleic acid (PNA) peptide conjugates and locked nucleic acid (LNA) oligomers for specific reduction of mutant huntingtin expression [42]. PNAs are a class of DNA/RNA mimic with an uncharged amide backbone, whose advantage is the lack of charge-charge repulsion of RNA/DNA or RNA/RNA duplexes, when these agents bind complementary RNA sequences. In this way, there is an increase on the affinity of PNA hybridization and the recognition of target sequences. On the other hand, LNAs are RNA analogs containing a methylene bridge between the 2'-oxygen and 4'-carbon of the ribose. This bridge reduces the conformational flexibility of ribose and confers outstanding affinity to complementary hybridization. These scientists observed that these two strategies can confer potent allele-selective inhibition of mutant Huntington gene expression [42].

Another strategy recently tested was based on the hypothesis that the epigenetic modifications close to GAA repeats present in Friedreich's ataxia can be reversed by pharmacological modulation during somatic cell reprogramming [43]. In this work, researchers reprogrammed Friedreich's ataxia fibroblasts into induced pluripotent stem cells in the presence of various small molecules that target DNA methylation and histone acetylation and methylation. Treatment of these stem cells with two compounds (sodium butyrate and Parnate), led to an increase in expression of the frataxin gene which persisted for several passages [43]. However, they observed that prolonged culture of those epigenetically modified cells resulted in progressive expansion of triplet repeats and consequent decrease of gene expression.

CONCLUSION

Over 3 billion bp of DNA constitute the human genome and a considerable part is formed by repetitive DNA [44]. This repetitive DNA includes microsatellites, telomere repetitive DNA sequences, centromeres and regions of chromosome heterogeneous DNA [44].

It is quite frequent the appearance of repetitive DNA that consists of repetition of triplets that can be present in several genes and code several aminoacids [44]. These triplet repeats can appear in enzymes, hormone receptors or non-coding regulatory sequences. These repeats can disrupt protein structure or interfere with cellular activities like methylation and imprinting, protein degradation through ubiquitination, and processing through proteasome, as well as gene activation and inactivation [44].

Triplet repeats can appear during somatic or germline cell division as well as in non-dividing cells [22]. The instability of triplet repeats can lead to gradual expansion of the repeat that might result in disease [44]. Expansion of these repetitions is particularly important in neurons and has an important role in brain aging, as well as in the appearance and progression of diseases in mice and humans [22].

Expansion of triplet repeats was first identified as happening during DNA replication, however, studies using bacteria, yeasts and mice showed that it can also occur during DNA repair, transcription and in recombination [22]. However, mechanisms involved in expansion are still only slightly understood [22].

There are even cases of bigger expansion or contraction of triplet repeats depending on the kind of repeat and on its parental origin [44].

Expansion sizes can give rise to non-affected carriers (permutation) and, as the expansion size increases through following generations, can lead disease with increase in severity of clinical symptoms and age of onset [44].

Triplet repeat expansion over generations leads to the phenomenon of anticipation [44]. Studies in families with myotonic dystrophy or Huntington disease have shown strong evidences that anticipation has a biological and genetic basis in the form of the rise in triplet repeats number. This number increases in following generations showing an inverse correlation between the number of repeats and the age of symptoms onset [45].

In cases of genetic anticipation, being female or male influences disease severity. For example, in Huntington disease descendants of affected males have an earlier onset of the disease when compared with descendants of affected women [45].

Most triplet repeats diseases have parental origin, since it is very rare the occurrence of de novo mutations [44]. However, the existence of specific DNA replication and repair mechanisms having the goal of detecting mistakes and repairing DNA, can potentially explain why some persons are more likely to suffer expansions or contractions of triplet repeats [44].

Several treatment strategies have been developed for this group of disorders, some directed to triplet repeat in DNA and others aiming the transcribed RNA. Nowadays, the main concern is to target mutant gene expression, without damaging transcription of the wildtype gene.

REFERENCES

1. Boeva V, Regnier M, Papatsenko D, et al. Short fuzzy tandem repeats in genomic sequences, identification, and possible role in regulation of gene expression. Bioinformatics 2006;22:676-84.
2. Huntley MA, Golding GB. Neurological proteins are not enriched for repetitive sequences. Genetics 2004;166:1141-54.
3. Subramanian S, Madgula V, George R, et al. Triplet repeats in human genome: distribution and their association with genes and other genomic regions. Bioinformatics 2003;19:549-52.
4. McConnell R, Middlemist S, Scala C, et al. An unusually low microsatellite mutation rate in Dictyostelium discoideum, an organism with unusually abundant microsatellites. Genetics 2007;177:1499-507.

5. McMurray CT. Mechanisms of trinucleotide repeat instability during human development. Nat Rev Genet 2010;11:786-99.
6. Krzyzosiak WJ, Sobczak K, Wojciechowska M, et al. Triplet repeat RNA structure and its role as pathogenic agent and therapeutic target. Nucleic Acids Res 2012;40:11-26.
7. Sobczak K, Michlewski G, Mezer M, et al. Structural diversity of triplet repeat RNAs. J Biol Chem 2010;285:12755-64.
8. Jakupciak JP, Wells RD. Genetic instabilities in (CTG.CAG) repeats occur by recombination. J Biol Chem 1999;274:23468-79.
9. Jung J, van Jaarsveld MT, Shieh SY, et al. Defining genetic factors that modulate intergenerational CAG repeat instability in Drosophila melanogaster. Genetics 2011;187:61-71.
10. Jakupciak JP, Wells RD. Genetic instabilities of triplet repeat sequences by recombination. IUBMB Life 2000;50:355-9.
11. Long A, Napierala JS, Polak U, et al. Somatic instability of the expanded GAA repeats in Friedreich's ataxia. PLoS One 2017;12.
12. Deighton C, Criswell LA, Lum RF, et al. Ages of onset suggestive of genetic anticipation in rheumatoid arthritis multicase sibships can be explained by observational bias. Rheumatology 2007;46:120-3.
13. Lima IVM. Repetições CAG: candidatos na génese das psicoses funcionais. Rev Bras Psiquiatr 1999;21:23-6.
14. Shekari Khaniani M, Aob P, Ranjouri M, et al. Molecular analysis and prevalence of Huntington disease in northwestern Iran. Turk J Med Sci 2017;47:1880-4.
15. Young E, Sloan J, Riper K. Trinucleotide repeats are clustered in regulatory genes in Saccharomyces cerevisiae. Genetics 2000;154:1053-68.
16. Calabrese PP, Durrett RT, Aquadro CF. Dynamics of microsatellite divergence under stepwise mutation and proportional slippage/point mutation models. Genetics 2001;159:839-52.
17. Rolfsmeier M, Dixon M, Pessoa-Brandão L, et al. Cis-elements governing trinucleotide repeat instability in Saccharomyces cerevisiae. Genetics 2001;157:1569-79.
18. Chen JL, Fountain MA, Yildirim I, et al. Structure and dynamics of RNA repeat expansions that cause Huntington's disease and Myotonic Dystrophy type 1. Biochemistry 2017;56:3463-74.
19. Stevens J, Lahue E, Li G, et al. Trinucleotide repeat expansions catalyzed by human cell-free extracts. Cell Research 2013;23:565-72.
20. Pinto RM, Dragileva E, Kirby A, et al. Mismatch repair genes Mlh1 and Mlh3 modify CAG instability in Huntington's disease mice: genome-wide and candidate approaches. PLoS Genet 2013;9.
21. Liu Y, Wilson SH. DNA base excision repair: a mechanism of trinucleotide repeat expansion. Trends Biochem Sci 2012;37:162-72.
22. Everett CM, Wood NW. Trinucleotide repeats and neurodegenerative disease. Brain 2004;127:2385-405.
23. Hou M-H, Robinson H, Gao Y, et al. Crystal structure of actinomycin D bound to the CTG triplet repeat sequences linked to neurological diseases. Nucleic Acids Res 2002;30:4910-7.
24. La Spada AR, Weydt P, Pineda VV, et al. Frontiers in neuroscience Huntington's disease pathogenesis: mechanisms and pathways. In: Neurobiology of Huntington's disease: Applications to Drug Discovery; Lo DC, Hughes RE. 2011.
25. Muragaki Y, Mundlos S, Upton J, et al. Altered growth and branching patterns in synpolydactyly caused by mutations in hoxd13. Science 1996;272:548-51.
26. Cohen MMJr. Holoprosencephaly: clinical, anatomic, and molecular dimensions. Birth Defects Res. Part A Clin Mol Teratol 2006;76:658-73.
27. Shibata A, Machida J, Yamaguchi S, et al. Characterisation of novel runx2 mutation with alanine tract expansion from Japanese cleidocranial dysplasia patient. Mutagenesis 2016;31:61-7.
28. Goodwin M, Mohan A, Batra R, et al. MBNL sequestration by toxic RNAs and RNA misprocessing in the Myotonic Dystrophy brain. Cell Rep 2015;12:1159-68.
29. Boivin M, Willemsen R, Hukema RK, et al. Potential pathogenic mechanisms underlying Fragile X Tremor Ataxia Syndrome: RAN translation and/or RNA gain-of-function? Eur J Med Gen 2017; S1769-7212(17):30396-8.
30. Glineburg MR, Todd PK, Charlet-Berguerand N, Sellier C. Repeat-associated non-AUG (RAN) translation and other molecular mechanisms in Fragile X Tremor Ataxia Syndrome. Brain Res 2018:300064-7.
31. Bañez-Coronel M, Aythan F, Tarabochia AD, et al. RAN translation in Huntington disease. Neuron 2015;88:667-77.
32. Hutvagner G, Simard MJ. Argonaute proteins: key players in RNA silencing. Nat Rev Mol Cell Biol 2008;9:22-32.
33. Wild EJ, Tabrizi SJ. Therapies targeting DNA and RNA in Huntington's disease. Lancet Neurol 2017;16:837-47.
34. Kordasiewicz HB, Stanek LM, Wancewicz EV, et al. Sustained therapeutic reversal of Huntington's disease by transient repression of huntingtin synthesis. Neuron 2012;74:1031-44.
35. Naryshkin NA, Weetall M, Dakka A, et al. Motor neuron disease. SMN2 splicing modifiers improve motor function and longevity in mice with spinal muscular atrophy. Science 2014;345:688-93.
36. Agustin-Pavón C, Mielcarek M, Isalan M, et al. Deimmunization for gene therapy: host matching of synthetic zinc finger constructs enables long-term mutant huntingtin repression in mice. Mol Neurodegener 2016;11:64.
37. Provenzano C, Cappella M, Valaperta R, et al. CRISPR/Cas9-mediated deletion of CTG expansions recovers normal phenotype in myogenic cells derived from Myotonic Dystrophy 1 patients. Mol Ther Nocleic Acids 2017;9:337-48.
38. Shin JW, Kim KH, Chao MJ, et al. Permanent inactivation of Huntington's disease mutation by personalized allele-specific CRISPR/Cas9. Hum Mol Genet 2016;25:4566-76.
39. Li L, Shen X, Liu Z, et al. Activation of frataxin protein expression by antisense oligonucleotides targeting the mutant expanded repeat. Nucleic Acid Ther 2018;28:23-33.
40. Corey DR. Nusinersen, an antisense oligonucleotide drug for spinal muscular atrophy. Nat Neurosci 2017;20:497-9.
41. Hu J, Matsui M, Corey DR. Allele-selective inhibition of mutant huntingtin by peptide nucleic acid-peptide conjugates, locked nucleic acid and small interfering RNA. Ann N Y Acad Sci 2009;1175:24-31.
42. Polak U, Li Y, Butler JS, et al. Alleviating GAA repeat induced transcriptional silencing of the Friedreich's ataxia gene during somatic cell reprogramming. Stem Cells Dev 2016;25:1788-800.
43. Lutz RE. Trinucleotide repeat disorders. Semin Pediatr Neurol 2007;14:26-33.
44. Polito J, Rees R, Childs B, et al. Preliminary evidence for genetic anticipation in Crohn's disease. Lancet 1996;347:798-800.