

Familles atypiques et rôle des interruptions sur l'instabilité des triplets CTG impliqués dans la Dystrophie Myotonique de type 1 (DM1)

Elodie Dandelot

► To cite this version:

Elodie Dandelot. Familles atypiques et rôle des interruptions sur l'instabilité des triplets CTG impliqués dans la Dystrophie Myotonique de type 1 (DM1). Génétique. Université Paris Cité, 2019. Français. NNT : 2019UNIP5089 . tel-04008729

HAL Id: tel-04008729 https://hal.sorbonne-universite.fr/tel-04008729v1

Submitted on 25 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Université de Paris

BioSPC

Laboratoire CTG-DM / Inserm U1163 et UMRS 974

Familles atypiques et rôle des interruptions sur l'instabilité des triplets CTG impliqués dans la Dystrophie Myotonique de type 1 (DM1)

Par Elodie DANDELOT

Thèse de doctorat de Génétique

Dirigée par Dr. Geneviève GOURDON

Présentée et soutenue publiquement le 28 Octobre 2019

Devant un jury composé de :

MERIENNE Karine (DR)

DION Vincent (Professeur)

JAGLA Krysztof (DR)

Rapporteur

Examinateur

Rapporteur

VIDAUD Michel (Professeur) Examinateur

GOURDON Geneviève (DR) Directeur de thèse

• Except where otherwise noted, this work is licensed under http://creativecommons.org/licenses/by-nc-nd/3.0/

Titre : Familles atypiques et rôle des interruptions sur l'instabilité des triplets CTG impliqués dans la Dystrophie Myotonique de type 1 (DM1)

Résumé :

L'instabilité des répétitions de microsatellites est un facteur clé pour plus de 40 maladies humaines. Parmi ces maladies, la Dystrophie Myotonique de type 1 (DM1) est une maladie neuromusculaire autosomique dominante caractérisée par un fort phénomène d'anticipation : les symptômes s'aggravent et apparaissent plus tôt d'une génération à l'autre. La DM1 est causée par une expansion de triplet CTG répétés dans la région 3'UTR du gène DMPK. Dans la population générale. la répétition CTG est stable et varie de 5 à 37 répétitions. Les patients DM1 ont des répétitions variant de 50 à > 1 000 CTG et sont instables d'une génération à l'autre (instabilité intergénérationnelle) et dans les tissus (instabilité somatique), avec un biais vers les expansions. En général, il existe une bonne corrélation entre la longueur des répétitions anormales et la gravité des symptômes. L'instabilité des triplets CTG implique divers facteurs (réparation, transcription, réplication, structures secondaires...). Cependant, les mécanismes de contractions des triplets CTG restent peu compris. Des interruptions dans les répétitions ont été rapportées chez des patients présentant une instabilité des triplets CTG réduite. Il est suggéré que les interruptions pourraient être un facteur de stabilisation ou de contraction mais jusqu'à présent, aucune démonstration directe de leur impact n'a été présentée. L'objectif de ma thèse est de démontrer si les interruptions impactent l'instabilité des triplets CTG et par quels mécanismes.

Nous avons étudié des familles DM1 atypiques, exemptes de phénomène d'anticipation et présentant des contractions de triplets CTG sur 2 à 4 transmissions successives. Tous les patients de ces familles présentent des interruptions : un seul CAG en 5' de la répétition CTG (famille A), ou plusieurs interruptions de la CCG en 5' ou 3' de la répétition (dans les familles B et E respectivement). Nous avons démontré que l'amplitude de l'instabilité somatique est réduite au moins pour les familles A et B. Ces d'interruptions sont de bons candidats pour étudier les mécanismes de contraction/ stabilisation des répétitions CTG.

J'ai tout d'abord démontré in vitro que les structures secondaires de l'ADN étaient différentes entre les répétitions pures ou interrompues. J'ai ensuite comparé la fixation des protéines nucléaires aux répétitions CTG, avec ou sans interruption, par EMSA (Electrophoretic Mobility Shift Assay). Mes résultats ont démontré des profils de fixation différents selon la nature des répétitions. Afin d'étudier plus précisément l'impact des interruptions sur l'instabilité des répétitions, j'ai dérivé des modèles de cellules HEK293 en utilisant la technologie Phi-Integrase permettant l'intégration de plasmides d'expression. Ces cellules permettent la transcription bidirectionnelle de la région 3 'UTR du gène DMPK mutant portant des répétitions CTG pures ou interrompues avec une unique interruption CAG en 5' des répétitions CTG, reproduisant le motif identifié chez la famille A. Grâce à ce modèle cellulaire, j'ai démontré que la mosaïque somatique des répétitions CTG interrompues était réduite comparée à celle présentée par les clones à répétitions pures reproduisant ainsi ce qui a été précédemment observée chez la famille A comparée à des contrôles DM1 à répétitions pures. En utilisant ce modèle cellulaire, j'ai montré que la formation d'hybrides ADN/ARN (Rloops) est réduite dans les cellules avec une seule interruption CAG et qu'il y a une bonne corrélation entre la formation d'R-loop et le niveau de transcription de DMPK mutant antisens. Ainsi, ce modèle cellulaire démontre que l'interruption de CAG a un impact sur l'instabilité des répétitions CTG, impliquant notamment le métabolisme des R-loops.

Une meilleure compréhension des mécanismes de contraction ou de stabilisation des répétitions anormales de CTG est un point crucial pour pouvoir guérir le défaut moléculaire responsable de la DM1. Mon projet fournit de nouvelles informations sur les mécanismes de l'instabilité des triplets CTG, qui peuvent fournir de nouvelles cibles thérapeutiques.

Mots clefs : DM1, répétitions CTG, instabilité de triplets répétés, interruptions, R-loop, structure secondaire de l'ADN, interaction ADN et protéines

Title : Atypical Myotonic Dystrophy type 1 families and how interruption(s) impact CTG repeat instability.

Abstract :

The instability of microsatellite repeats is a key factor for more than 40 human diseases. Among these diseases, Myotonic Dystrophy Type 1 (DM1) is an autonomic dominant neuromuscular disease characterized by a strong phenomenon of anticipation: the symptoms worsen and appear earlier from one generation to the next. DM1 is caused by expanded CTG repeats in the 3'UTR of the *DMPK* gene. In the general population, the CTG repeat is stable and ranges from 5 to 37 repetitions. Patients with DM1 have repeats that range from 50 to> 1000 CTG and are unstable from one generation to the next (intergenerational instability) and in the tissues (somatic instability), with a bias towards expansions. In general, there is a good correlation between the length of abnormal repeats and the symptoms severity. Instability of the CTG repeats involves various factors (DNA repair, transcription, replication, secondary structures etc...). However, the CTG repeats contraction mechanisms remain poorly understood. Repeat interruptions have been reported in patients with reduced CTG instability. It is suggested that interruptions could be a stabilization or contraction factor but so far, no direct demonstration of their impact has been presented.

The aim of my thesis is to demonstrate whether and how interruptions impact the instability of CTG repeat expansions.

We studied atypical DM1 families, with no anticipation and associated with contractions of the repeats on 2 to 4 successive transmissions. All patients in these families showed interruptions: a single CAG in 5 'of the CTG repeats (family A), or several CCG interruptions in 5' or 3 'of the repeat (in families B and E respectively). We demonstrated that the amplitude of somatic instability was reduced at least for families A and B. These interruptions are good candidates to learn more about the contraction / stabilization mechanisms of CTG repeats.

I first demonstrated in vitro that DNA secondary structures were different between pure or interrupted CTG repeats. Then, I compared the binding of nuclear proteins to CTG repeats with or without interruption by EMSA (Electrophoretic Mobility Shift Assay). My results demonstrated different binding profiles depending on the nature of the repeats. To further study the impact of interruptions on CTG instability, I derived HEK293 cell models using Phi-Integrase technology and plasmids expressing the CTG repeats. These cells allow bidirectional transcription of the 3' UTR region of mutant DMPK carrying pure or CTG repeats with a single CAG interruption in 5' of the repeats, as the interruption pattern observed in family A.In these cells, I demonstrated via the newly optimized Flash-Small-Pool-PCR that the somatic mosaic of interrupted CTG repeats was reduced compared to that observed for pure repeat. I have also detected that the formation of DNA / RNA hybrids (R-loop) is reduced in cells with the single CAG interruption compared to the clones with pure repeats. Moreover, there is a good correlation between the R-loops level and the transcriptional pattern of mutant DMPK antisense transcripts. Thus, our cellular model demonstrates that the CAG interruption has an impact on the instability of the CTG repeats, implying at least the metabolism of the Rloops.

A better understanding of the mechanisms of contraction or stabilization of abnormal CTG repeats is crucial to being able to cure the molecular defect responsible for DM1. My project gives new insights into the mechanisms of instability of CTG repeats, which can provide new therapeutic targets and new clues to improve the prognosis of DM1.

Keywords : DM1, CTG repeats, triplet repeats instability, interruptions, R-loops, DNA secondary structures, DNA binding, DM1 human cell model

A Suzy, Cécile et Annette qui, toutes les trois à leur manière, m'ont inspirée une volonté de « faire ».

REMERCIEMENTS

« Les applaudissements ne se mangent pas. » (Maguy Marin) Oui mais ça me fait plaisir de les communiquer alors on va quand même faire plein de pages avec de bons sentiments et rendre justice à des gens bien !

Ça y est, la thèse est enfin rédigée, les expériences certainement pas terminées et l'addiction au café bien trop avancée... mais ça y est, le temps des remerciements est arrivé. C'est le moment de ne pas se tromper, c'est en général l'un des trucs les plus lus dans ce genre de pavé. J'en sais quelque chose, je les lis toujours, c'est édifiant. Alors à vous qui lisez ces pages entre deux cafés (ou thé, chacun sa religion et si la vôtre n'est pas encore à base d'alcaloïdes légaux, n'ayez pas peur, il n'y a pas d'âge pour une conversion en règle...) armez-vous de patience, car bien que je craigne de ne pas pouvoir être exhaustive, l'expression « travail d'équipe » n'a jamais été aussi vraie que pour cette thèse, et j'adoooooore les parenthèses difficilement fermées et surtout à base d'envolées lyriques à la syntaxe plus que douteuse. Bien entendu, je remercie mes rapporteurs Vincent Dion et Karine Merienne d'avoir accepté d'être rapporteurs pour cette thèse.

Et tout naturellement, je commence par vous Geneviève... (les 3 petits points c'est pour donner un petit suspens). MERCI ! En lettres majuscules parce que ça rend à peine justice à ma gratitude envers vous. Jamais je ne pourrai vous remercier assez (parce que sinon il y aurait sans doute une gêne légitime qui se créerait, puis une ordonnance d'éloignement...bref beaucoup trop de paperasse dont, je suis sûre, vous préfèreriez vous passer). Sans vous, et cela à bien des égards, jamais cette thèse n'aurait pu se faire. J'ai eu la chance de pouvoir compter sur vous, sur votre patience (si, si...), sur votre bienveillance constante, votre amour contagieux des sciences, votre exigence aussi flatteuse qu'agaçante par manque de sommeil. Merci pour la confiance que vous m'avez conférée. Je ne dis pas que je n'ai jamais pesté après vous (de toutes façons vous ne le croiriez pas, vu que je peste à peu près tout le temps sur tout et n'importe quoi). J'ai conscience de la chance que j'ai eu de vous avoir à mes côtés dans ce monde parfois un peu tordu qu'est le doctorat. Merci encore pour tout, vous avoir en directrice de thèse a été une chance inestimable. Par je ne sais quelle magie, vous avez réussi à me redresser face à de nombreux tumultes, à m'insuffler confiance en ma légitimité en tant que thésarde (et ce n'était pas gagné !). Alors pour la -énième fois, merci infiniment et surtout je vous souhaite une bonne continuation dans la nouvelle aventure REDS (mais je ne me fais aucun souci !). Je pourrais continuer mon discours sur vous, et vous le mériteriez, mais les suivants seraient jaloux...

L'instant est délicat, qui faire suivre à présent ?

Petite pause pour le suspens n°2...

Mario ! Merci pour avoir supporté les réunions juuuuuuste à côté de ton bureau, merci pour avoir développé ma culture de l'Eurovision, et de m'avoir mis en lumière bon nombre d'incongruités françaises (le « bonne fin d'appétit » restera inoubliable). Ton imitation de danse de la poule est aussi un collector dans mes souvenirs... Merci aussi et surtout pour toutes les remarques et idées constructives que tu m'as transmis. Aline, mais que ferais-je sans toi ? Ma première Small-Pool c'était avec toi, et tu m'avais pourtant prévenue : cette manip c'est l'enfer. Et entre 2 buffers approximatifs, tu avais parfaitement raison. Comme toujours? Toi qui me mets face à mes mots, mes manies pour mieux en rire, toi qui as toujours été là pour me faire relativiser et rationaliser les choses (ok t'étais pas la seule, mais tu le fais super bien !). Toi qui étais aux premières loges pour mes seules et traumatiques dissections de DMSXL, et qui m'a fait découvrir que le Bourgogne a de bons arguments de vente. Toi qui nous as accueillis comme des rois pour des aventures épiques en canoë, toi qui me rappelais à l'ordre quand je défrisais à la moindre manip ratée... j'en oublie des choses et sûrement des meilleures, mais merci pour tout. Parce que même si (je te cite) « si je t'avais croisée dans la rue, jamais je ne t'aurai adressé la parole, on ne serait clairement pas copines ... » et bien heureusement la vie de labo est passée par là, ça aurait été dommage de ne pas ajouter ton personnage à mes connaissances. (Avoue que t'as été curieuse de les lire ces remerciements...ça va les chevilles ? Ce long paragraphe, c'est pour rattraper ma tendance à te faire faire 1m10 sur les dessins d'équipe...). Hélène, membre du duo des mamans du bureau. Tu nous as rejoint un peu plus tard, mais t'avoir en voisine de paillasse a été l'objet de bon nombre de discussions parfois sérieuses, parfois amusantes, toujours agréables. Une voisine de paillasse qui patiemment écoutais mes traits d'esprits pas toujours très inspirés pour commenter le contenu de la radio et qui ne peste presque pas alors que je fiche le chantier quotidiennement, c'est merveilleux... Sandra, my dear Ph.D fellow/ life saviour...mais comment trouver les mots pour résumer tous ces regards de jugement sur mon café allongé accompagné d'une tranche de brioche avec du chocolat? Comment reporter avec justesse tous ces post-it, tous ces rires "Under pressure", et le nombre incalculable de fois où tu as été là pour me remonter le moral ou me secouer pour me redresser ? Ta présence et sagacité ont été des trésors précieux. Je suis chanceuse d'avoir eu une compagne de galère comme toi. J'espère qu'on se croisera à nouveau, en blouse ou avec un Gin en terrasse entre Paris et Porto. En tout cas, comptes que tu as ici une admiratrice ! Diana, Louison, duo de génies au karma à dimensions variables, duo implacable à la force de travail contagieuse et la bonne humeur communicative. Je suis décidément quelqu'un de chanceux d'avoir pu travailler avec vous dès le début en M2. Piliers de la « Best Team ever » à l'addiction aux sushis assumée, et propagandistes healthy de choc, rien n'aurait été pareil sans vous. Tous ces weekends et ces soirées, au labo, au café ou ailleurs sont des pépites de réconfort et de sourires. Des personnes comme vous, on n'en croise pas si souvent, je suis fière d'avoir travaillé à vos côtés et de vous compter parmi mon petit univers, et Louison, je m'excuse pour tous ces traumas à bases de chaussettes dépareillées...

Les stagiaires : merci pour votre totale dévotion et reconnaissance de notre suprématie à Diana, Sandra, Louison et moi-même. Quoi vous vous attendiez à mieux ?! Sérieusement ?! Bah Alexis merci pour les 153kg de bonbons rapportés religieusement, Antoine pour les discussions édifiantes sur la supériorité de Batman sur l'univers Marvel, et Aurélien pour être un troll de compétition avec qui le sarcasme et l'ironie sont des sports de compétition. Bon, sérieusement (ou pas d'ailleurs), merci, c'était bien sympa de vous tolérer dans le bureau, et surtout, bonne continuation à vous, bien que je conçoive que l'idée de ne plus être dans le même bureau que Diana, Sandra, Louison, Aline, Hélène et moi soit difficile à supporter. Merci. Stéphanie de m'avoir démontré la force de l'auto-détermination. Il en faut, et je dois avouer que cette notion me manquait sûrement un peu avant de te croiser. Notre rencontre est une expérience qui fait date dans mon parcours, et je ne l'oublierai jamais (enfin, comme tu le sais, j'ai une mémoire de poisson rouge en état de sévère déshydratation, cette affirmation est donc à prendre avec des pincettes). Mouly and Anchel, thanks for your kindness, and your humor which will always stay with me.

Olivier, Tristan, Daniel, Cyril, noyau dur d'un petit groupe dont il est difficile de repousser une invitation en terrasse... toutes ces soirées à refaire le monde, pester, ironiser et parfois à ne rien comprendre à ce que vous pouvez raconter, les cafés de la qualité, les échanges dans les couloirs, et l'entraide, votre amitié... tout ceci constituera une partie inaliénable de ma thèse, merci d'avoir été là aussi ! Evelyne, il faut que je te parle, je pense que Quinzel et Thor fomentent un plan d'asservissement des humains devant de déployer sous peu...mais surtout merci pour ces lettres et ces mots outre-Atlantique, promis un jour on brûlera plein de trucs ensembles. Christine, j'ai toujours pu trouver ton bureau ouvert, je me suis toujours sentie accueillie autour d'un yogi-tea, toujours un mot pour redresser la barre : merci pour tout. Merci aux « voisins Antignac-Saunier», ceux qui sont déjà partis et dont je ne suis pas prête d'oublier les personnalités, je parle de vous Gweltas, Rebecc', Albane, Vale... et merci aussi aux actuels qui m'ont permis de trouver un petit havre de paix pour cette délicate période qu'est la rédaction. Merci aussi à Amine et Moussa, toujours prompts à me faire rire, à prendre des nouvelles, de vrais phares dans cet institut aux murs trop blancs et ascenseurs défaillants.

J'en oublie sûrement, mais j'imagine le lecteur qui s'attarde sur les remerciements et qui se dit : ça n'en finira donc jamais ? Si, ça se finira sûrement, mais cette thèse a été avant tout un parcours humain, et bizarrement on ne peut pas trop le caser dans l'intro ou le Matériels et Méthodes...c'est un tort.

Papa, maman, vous le savez, d'ordinaire, je suis quelqu'un qui parle beaucoup, voire trop. Et pourtant j'en viens à ne pas savoir quoi écrire pour vous témoigner toute ma reconnaissance, toute mon affection à votre égard. J'ai la chance de vous avoir avec moi, d'avoir toujours pu compter sur vous, depuis toujours. Vous m'avez appris à être qui je suis aujourd'hui, à avoir envie de faire. Merci de m'avoir supportée dans mes études et dans la vie avec toute votre affection qui est une source inépuisable de force. Lucile, Margot, je ne compte pas le nombre de fois où j'ai pesté avec vous, où je vous ai trainées au labo alors que vous passiez sur Paris, sans jamais me faire comprendre que j'abusais de vous emmener en chambre froide en plein mois de février. Vous avez continuellement été d'un réconfort indispensable, j'espère sincèrement l'être un peu pour vous aussi. Noémie, chère cousine toujours prête à m'emmener voir des univers merveilleux sur scène et surtout pour partager un bon moment. Combien de fois t'ai-je posé des lapins (et ce n'étaient pas Quinzel ou Jiminy) sans que jamais tu n'en prennes ombrage ? Ton parcours est un modèle de détermination et de courage, derrière mes pipettes, j'ai pu te voir te réaliser dans ton parcours, et c'est chouette. Papi, Annette, Suzy, Cécile, chacun à votre manière avez contribué humainement à cette thèse et j'en suis aussi fière qu'heureuse : vous avez soutenu un petit troustet cap burut (je n'ai aucune idée de l'orthographe de ces mots) depuis longtemps, et aujourd'hui encore j'apprends de vous. Géno, tous ces moments de partages, ces heures à m'écouter, me réconforter sont inestimables, tu es une présence tranquillisante en ces territoires France-îliens, merci à toi aussi pour tout, ces dernières années ou même avant : si toi qui lis ces remerciements tu me connais un peu, sache que Géno est à la base de mon coup de crayon...et oui ! Merci à Pascale et Bernard pour ces soirées princières, ces pauses avec vous ont aussi grandement contribué à l'accomplissement de cette thèse. Cathy, Jean-Marie, Frédé, merci à vous aussi pour toutes ces fois où, de passage sur Paris, vous aviez toujours un moment à passer avec moi et mon emploi du temps impossible.

Merci à Jeannette Pétrolette, Anaïs, Ariane, Mathilde, Justine, Clara, Noémie, ladies que j'ai la chance de pouvoir compter parmi mes amies. Ladies aux multiples qualités qu'il serait trop long d'énumérer. Amies de presque toujours ou amitiés plus parisiennes, toutes arrivant avec leurs horizons divers et leur soutien inébranlable : autour d'un verre, d'une table de JDR, d'une conversation whatsapp de quelques messages ou de costumes des années 20 dans l'hôtel Shinning de Duddley... vous êtes trop géniales mesdames.

Et enfin, merci à un certain papillonneur aux yeux bleus qui, par je ne sais quelle alchimie et autres trésors d'affection, est une étincelle inaltérable qui m'illumine d'un peu plus de tendresse chaque jour, avec une patience d'or.

(oui c'est fini là, pour de bon, mais j'aurais pu continuer !)

LISTE DES PRINCIPALES ABREVIATIONS

- ADN : Acide Désoxyribonucléique
- ARN: Acide Ribonucléique
- **TNRs:** Trinucleotides Repeats
- UTR: Untranslated Region
- DM1: Dystrophie Myotonique de type 1
- HD: Huntington Disease
- FSX: Syndrome de l'X Fragile
- CTG: Cytosine, Thymine, Guanine
- DMPK: Dystrophia Myotonica Protein Kinase
- Pb: Paire de Bases
- Kpb : Kilo paires de bases
- POLR2A: ARN polymérase II ADN-dépendante
- MMR : Mécanisme de Réparation des Mésappariements de bases
- BER : Réparation par Excision de Base
- NER : Réparation par Excision de Nucléotide
- IP : Immuno-Précipité

SOMMAIRE

REMERCIEMENTS	4
LISTE DES PRINCIPALES ABREVIATIONS	8
TABLE DES TABLEAUX	
Table des Figures	13
INTRODUCTION	1
I. Les sequences repetees dans le genome humain	2
I.1 Les différents types de séquences d'adn répétées	2
I.1.1. Les séquences répétées dispersées	2
I.1.1. Les séquences répétées en tandem	5
I.2. Les séquences microsatellites dans le génome humain	6
I.2.1. Répartition au sein du génome	6
I.2.2. Les répétitions trinucléotides dans le génome humain	7
II. Les maladies humaines a repetitions de nucleotides	8
II.1. Aperçu général	8
II.2. Pathologies causées par des triplets répétés dans les exons	13
II.2.1. Maladies à expansion d'alanine (GCG)	
II.2.2. Maladies à expansion de glutamines (Poly-Q)	
II.3. Pathologies causées par des triplets répétés dans des régions non codantes des gènes	
II.3.1 Expansions CGG/CCG : Syndrome de l'X fragile et maladies associées	
II.3.2. Maladie à expansions gaa : l'ataxie de friedreich (frdra ou fa)	
II.3.3 Maladie à expansion cag/ctg – sca12	22
II.3.4 Maladies à expansions ctg	23
II.3. Conclusion de chapitre	32
III- INSTABILITE DES TRIPLETS REPETES	
III-1. Instabilité intergénérationnelle	33
III.1.1. Influence du parent transmetteur : évènements germinaux	
III-1.2. Instabilité au cours du développement : Évènements zygotiques	
III.2. Instabilité somatique	
III.3. Facteurs d'influence de l'instabilité	
III.3.1. La pureté des répétitions	
III.3.2. La longueur des répétitions	
III.3.3. Les structures secondaires	
III.3.4. Les séquences environnantes	44
III.4. Métabolismes de l'adn et instabilité	50
III.4.1. La Réplication	50
III.4.2. La réparation	59
III.4.3. La transcription	72

III.5. Conclusion de chapitre	82
IV. Strategies therapeutiques pour la dm1	83
IV.1. Approches médicamenteuses	83
IV.1.1. Diminution de l'expression de <i>dmpk</i> et conséquences	83
IV.1.2. mbnl1 hors des <i>foci</i>	84
IV.1.3. Régulation de la phosphorylation - impact sur celf1	
IV.2. Thérapie par l'arn	85
IV.3. Thérapies géniques	86
IV.4. Conclusion de chapitre : aperçu des essais cliniques	87
V. OBJECTIFS DE LA THESE	89
V-1. Familles dm1 atypiques et interruptions	89
V-2. Optimisation du protocole de Small-Pool-pcr	
V-3. Rôle(s) des interruptions dans l'instabilité des triplets ctg	91
RESULTATS	
I. FAMILLES DM1 ATYPIQUES ET INTERRUPTIONS	93
I.1. Éléments de contexte	
I.2. Résumé des principaux résultats	94
I.2.1. Identification d'interruptions	94
I.2.2. Comparaison de la mosaïque somatique des patientes à interruptions avec celle de patients dn	n1 contrôles 94
I.2.3. Analyse d'autres facteurs modulateurs de l'instabilité des tnrs	95
I.3. Conclusions générale et Perspectives	
I.4. Contribution	
I.5. Données non publiées	126
I.5.1 Quantification de l'expression sens et antisens de DMPK	126
I.5.2. Analyse de l'expression de Lig1	128
II. OPTIMISATION DU PROTOCOLE DE SMALL-POOL-PCR	130
I.1. Éléments de contexte	130
I.2. Résumé des principaux résultats	130
I.3. Conclusions générale et Perspectives	131
II.4. Contribution	131
THE FLASH-SMALL-POOL PCR PROTOCOL	136
PROTOCOL FOR:	136
III. ROLE(S) INTERRUPTIONS DANS L'INSTABILITE DES TRIPLETS CTG	144
III.1. Éléments de contexte	
III.2. Résumé des principaux résultats	145
III.2.1. Influence des interruptions sur la structure secondaire des répétitions	
III.2.2. Influence des interruptions sur la fixation des protéines nucléaires aux répétitions ctg	

III.2.3. Impact d'une seule interruption cag sur l'instabilité des triplets ctg	7
III.2.4. Impact d'une seule interruption cag sur le métabolisme des arn dmpk	8
III.3. Conclusions générale et Perspectives14	9
III.4. Contribution	9
III.5. Données non destinées à la publication17.	2
III.5.1. Impact d'une interruption unique cag sur l'instabilité des triplets ctg dans des cellules humaines non	
immortalisées17	2
III.5.2. Détection des R-loops sur fibroblastes primaires17	3
III.5.3 Impact des interruptions ccg sur le recrutement de proteines nucléaires sur les répétitions de ctg17	4
DISCUSSION ET PERSPECTIVES	6
I- FAMILLES DM1 ATYPIQUES ET INTERRUPTIONS	8
II- LA FLASH-SMALL-POOL PCR	0
III- ROLE DES INTERRUPTIONS DANS L'INSTABILITE DES TRIPLETS CTG	1
III.1. Une unique interruption cag diminue l'instabilité des triplets ctg	1
III.2. Une unique interruption cag influence la transcription de dmpk	2
III.3. Les interruptions modifient la structure secondaire des répétitions de ctg	3
III.4. Les interruptions modifient le recrutement de protéines nucléaires sur les répétitions de ctg	4
V. CONCLUSION GENERALE	5
BIBLIOGRAPHIE	7
ANNEXE 1 : TABLEAU RECAPITULATIF DES MOTIFS D'INTERRUPTION(S) DANS LES REPETITIONS CTG DE	
FAMILLES DE PATIENTS DM1	6
ANNEXE 2 : GENETIC MODIFIERS OF CAG.CTG REPEAT INSTABILITY IN HUNTINGTON'S DISEASE MOUSE MODEL	S
(REVIEW DANDELOT,E. AND TOME,S)23	0
ANNEXE 3 : RESUME ACTIVITES COMPLEMENTAIRES AU DOCTORAT	2

TABLE DES TABLEAUX

Tableau 1: Caractéristiques et exemples de séquences répétées en tandem dans l'ADN
humain5
Tableau 2: Aperçu des principales caracteristiques des pathologies dues a des repetitions
GCG dans des régions codantes (maladies polyalanines)10
Tableau 3: Caractéristiques des pathologies dues à une expansion CAG dans des régions non
codantes (poly-glutamine, polyQ)11
Tableau 4: Caractéristiques des pathologies dues à une expansion de triplets dans une région
non codante
Tableau 5: Exemples de protéines dérégulées dans la DM1 conséquences attribuées ou
auppación dens la physionathología de la DM1
Tableau 6: Protéines du BER impliquées dans l'instabilité des triplets dans des modèles
murins
Tableau 7: Exemples de démonstration de l'impact des protéines du MMR dans différents
modèles murins (1/2)67
Tableau 8: Exemples de démonstration de l'impact des protéines du MMR dans différents
madèles murine (2/2)
modeles munns (2/2)
Tableau 9: Ratio des gènes humains positifs pour la formation d'R-Loops selon leur classe.
Tableau 10: Exemples d'implication des R-loops dans les maladies à TNRs en cellules
humaines

TABLE DES FIGURES

Figure 1: Schématisation des Rétrotransposons présents chez l'Homme	4
Figure 2: Exemples de localisation génique et conséquences de différentes nature répétitions de nucléotides	s de 8
Figure 3: Observation et schématisation d'agrégats.	14
Figure 4: Exemple de famille atteinte de la Dystrophie Myotonique de Type 1	26
Figure 5: Représentation schématique de la pathogénèse de la DM1	28
Figure 6: Rappel sur les processus d'ovogénèse et spermatogénèse chez l'Homme	33
Figure 7: Schématisation d'hairpins composées de répétitions CTG.	40
Figure 8: Schématisation de différentes structure G-quadruplexes	42
Figure 9: Vue d'ensemble sur la nature des structures triplexes	43
Figure 10: Schématisation d'un DUE.	44
Figure 11: Détermination de la variation de la taille des triplets CTG en fonction positionnement de l'origine de réplication.	า du 46
Figure 12. : Méthylation en amont et en aval des répétitions CTG dans des cellules déri de patients DM1, CDM et de contrôles	vées 49
Figure 13: Schématisation des modèles "origin switch" et "origin shift"	51
Figure 14: Schématisation du modèle "Fork Shift"	52
Figure 15: Fourche de Réplication active	53
Figure 16: Arrêt de la fourche de réplication et différentes voies de dépassement de lé	sion. 56
Figure 17: Exemples de traitement des cassures doubles brin par la résolution des jonc d'Holliday	tions 63
Figure 18: Illustration du fonctionnement du MMR	65
Figure 19: Illustration de la stabilisation d'une hairpin par le complexe MSH2/MSH3	70
Figure 20: Aperçu général de la Transcription eucaryote	72
Figure 21: Schématisation d'une R-loop	76

igure 22: Schématisation de la formation co-transcriptionnelle d'R-loops dans les répétition	s
NG	0
igure 22: Approchae thérapoutiques visent à éliminer les fasi d'ADNm mutants DMDK	5
igure 23. Approches therapeutiques visant a entriner les foci d'ARIMIT Indiants DiviPRo	5
igure 24: Modèle intégratif de l'impact de l'interruption CAG dans l'instabilité des triplets CTG	i.
	6

INTRODUCTION

I. LES SEQUENCES REPETEES DANS LE GENOME HUMAIN

I.1 LES DIFFERENTS TYPES DE SEQUENCES D'ADN REPETEES

Selon les prédictions informatiques, les séquences répétées représenteraient entre 50% et 70% du génome humain *(de Koning et al. 2011)*. Ces séquences peuvent être grandement, moyennement ou légèrement répétées et sont impliquées dans les variations structurelles (SVs) du génome. Les SVs participent aux variations inter-individuelles *(Conrad et al. 2010)* et sont vecteurs de remodelage de l'ADN pouvant devenir pathologiques (maladies génétiques, cancers...). Il existe dans l'ADN humain 2 grands types de séquences répétées : (1) les séquences dites « dispersées », regroupant les éléments transposables, les gènes paralogues et les gènes codant l'ARN de transfert (ARNt) ; et (2) les séquences « répétées en tandem ».

I.1.1. LES SEQUENCES REPETEES DISPERSEES

Les séquences répétées dispersées représentent entre 40-45% du génome humain (*Prak et al. 2000, Pace et al. 2007*). Parmi ces séquences se trouvent les éléments transposables (transposons) qui se subdivisent en 2 grandes classes. La classe 1 des Rétrotransposons, la Classe 2 des Transposons à ADN. Les transposons à ADN sont très peu représentés chez l'humain (moins de 3%), et se déplacent selon le mode « couper/coller », c'est-à-dire sans système de réplication. Aucun transposon à ADN n'a été détecté comme « actif » chez l'Homme : nous ne les aborderons donc pas dans ce paragraphe. Nous aborderons cependant brièvement le cas des gènes paralogues, issus de copies de gènes ancestraux. Les différentes structures d'éléments transposables actives dans le génome Humain sont résumées dans la figure 1.

Les rétrotransposons sont issus de la rétro-transcription d'ARN initialement viral. Après rétro-transcription en ADNc, la séquence s'intègre au génome hôte. Les séquences insérées se multiplient ensuite au sein du génome selon différents modes.

Les rétrotransposons à LTR: ces éléments sont dits « autonomes » car les séquences LTR (Long Terminal Repeats) codent pour une rétrotranscriptase et une intégrase nécessaires à la multiplication de la séquence, ainsi qu'une protéase. Cette classe de rétrotransposons peut ainsi, en toute autonomie, produire des ADNc et les intégrer dans le génome hôte. Dans le génome humain, les rétrotransposons à LTR majoritaires sont les séquences HERVs, issues de rétrovirus endogènes, elles représentent 8% du génome (*Consortium 2001*).

Les rétrotransposons sans LTR : les principaux représentants de cette catégorie sont les séquences de type LINE (Long Interspesed Elements). Chez l'humain, la séquence LINE la plus représentée est « L1 » correspondant à 10% de l'ADN génomique (Consortium 2001). Les LINEs sont des éléments autonomes, car ils possèdent dans leur séquence une protéine unique, capable de les répliquer et de les intégrer dans l'ADN hôte. Dans cette classe sont aussi comprises les séquences type SINEs (Short INterspesed Elements), représentant 13% du génome. Les séquences SINEs comprennent les séquences Alu (5% du génome), les MER et les ARNt. Ces éléments ne produisent ni rétrotranscriptase, ni polymérase : ces séquences utiliseront la machinerie des rétrotransposons à LTR ou des LINE pour être mobiles dans le génome (*Klug 2009*).

Cependant, chez l'humain, quasiment toutes ces séquences en théorie « mobiles » restent stables et immobiles, à l'exception de certaines séquences L1 dont on estime que chaque individu porte entre 40 et 60 séquences « actives », donc mobiles. Les insertions *de novo* de L1 ont été cependant détectées et liées à des cas de Myopathie de Duchenne, bêta thalassémie et cancers (*Prak et al. 2000, Pace et al. 2007*).

Les gènes paralogues dérivent d'un même gène ancestral. Ce sont des copies homologues ayant divergé suite à des évènements de duplications de larges segments d'ADN. Les duplications inter-chromosomiques aboutissent à des répétitions paralogues dispersées. Les gènes paralogues chez l'être humain sont copiés sur une gamme allant de 5 à 368 copies à travers le génome. Certaines de ces duplications sont spécifiques à l'humain et sont étroitement liées au développement du cerveau et au phénomène de « conversion génétique » dans l'humanisation des primates (*Sudmant et al. 2010*).



Figure 1: Schématisation des Rétrotransposons présents chez l'Homme.

(A.) Les rétrotransposons LTR sont constitués de régions codantes qui se chevauchent partiellement. Ces séquences codent pour les gènes correspondant à des antigènes spécifiques du groupe (gag), une protéase (prt), une polymérase (pol) et enveloppe (env), flanquées des deux côtés par de longues répétitions terminales (LTR) avec une activité promotrice. Le gène Pol contient des domaines pour la transcriptase inverse (RT), la RNaseH et l'endonucléase (EN) nécessaires à la réplication/ré-intégration de séquences ADN. (B). Un élément L1 consiste en une région non traduite en 5' (5'-UTR) ayant une activité promotrice, deux cadres de lecture ouverts (ORF1 et ORF2) séparés par un spacer intergénique suivi d'une région 3'-UTR et d'une queue poly A (pA). L'ORF1 code pour une protéine qui se lie aux acides nucléiques, et l'ORF2 contient la transcriptase inverse (RT), un domaine endonucléase (EN) et une région riche en cystéine (C). (C.) Les SINEs, sont des séquences sans intron de 300 nucléotides composées de deux fragments riches en GC, le monomère gauche (L-Alu) et le monomère droit (R-Alu) sont reliés entre-eux par une séquence riche en Adénine, et qui se termine par une queue poly A (pA). L'élément est flangué de courtes répétitions directes (flèches), qui sont des duplications du site d'insertion génomique. (D.) Les SINE apparentés à l'ARN de transfert consistent en une région homologue à l'ARNt, une région non apparentée à l'ARNt (zone grise) et une queue polyA. La région en 3 'de certains SINE dérivés d'ARNt ressemble à une partie du 3' UTR des LINEs. D'Après (Prak and Kazazian Jr 2000). (E.) Répartition moyenne des différents éléments transposables dans le génome humain.

I.1.1. LES SEQUENCES REPETEES EN TANDEM

Les séquences répétées en tandem représentent 10% du génome humain. Ce sont des séquences de longueurs variables, qui se répétèrent les unes à la suite des autres. Ces répétitions regroupent les gènes paralogues répétés en tandems (issus de duplications intrachromosomiques), l'ADN ribosomique (ADNr) et les séquences satellites (*Richard et al. 2008*).

L'ADN ribosomique correspond à une séquence de 43kpb comprenant l'information génétique nécessaire pour construire les sous unités d'ARN ribosomique (ARNr) 18S, 5.8S et 28S et une séquence « spacer » intergénique. Ce motif de 43kpb se répète en tandem entre 200 et 600 fois sur le bras court de différents chromosomes (chromosomes 13,14,15,21 et 22) (*Agrawal et al. 2018*). Cette fréquence de répétition positionne l'ADN ribosomique dans la catégorie des séquences moyennement répétées chez l'humain (*Klug 2009*).

Les séquences satellites correspondent à des motifs plus ou moins courts, répétés en tandem. Les séquences satellites sont classées en fonction de la taille du motif répété, entre 1 et plusieurs centaines de nucléotides chez l'humain. Il existe ainsi les séquences mégasatellites, satellites, minisatellites et microsatellites : cette classification est décrite dans le tableau 1 (*Jeffreys et al. 1985, Kogi et al. 1997*). Leurs répartitions est inégale dans le génome, on retrouve par exemple les minisatellites principalement dans l'hétérochromatine comme les régions sub-telomériques et centrosomiques par exemple (*Charlesworth et al. 1994*). Dans la suite de ce chapitre, nous allons nous intéresser plus particulièrement aux séquences microsatellites.

	Taille de la séquence répétée	Longueur de la répétition	Exemples	Références
Mégasatellite	Plusieurs kb	Plusieurs centaines de kb	<i>Locus</i> RS447 (4p.15)	(Kogi <i>et al.</i> 1997)
Satellite	5-171pb	100kb à plusieurs Mb	Centromères	(Masumoto <i>et al.</i> 1989)
Minisatellite	6-100pb	500pb à 10kb	Télomères	(Charlesworth <i>et</i> <i>al.</i> 1994)
Microsatellite	1-10pb	<500pb	Répétitions (GGAT) dans le gène de la myoglobine	(Weller <i>et al.</i> 1984)

Tableau 1: Caractéristiques et exemples de séquences répétées en tandem dans l'ADN humain.

I.2. LES SEQUENCES MICROSATELLITES DANS LE GENOME HUMAIN

I.2.1. REPARTITION AU SEIN DU GENOME

La présence des séquences microsatellites, aussi appelées Short Tandem Repeat (STR) semble être corrélée à la complexité des génomes. En effet, les procaryotes n'ont que peu de microsatellites, et généralement les motifs s'y répètent peu *(Hancock 1996, Hancock 1996)*. Contrairement aux génomes eucaryotes qui présentent une forte densité de microsatellites : ils représentent 3% du génome humain par exemple. Les microsatellites sont uniformément répartis entre les chromosomes humain à raison d'une occurrence tous les 2Kpb en moyenne *(Consortium 2001)*, avec une densité légèrement plus élevée sur le chromosome 19 *(Subramanian et al. 2003)*. Les motifs rencontrés sont variables (entre 1 et 10pb), et le nombre de répétitions est polymorphes d'un individu à l'autre. Ces séquences sont notamment utiles dans la définition d'haplotypes *(Jeffreys et al. 1990)*.

Tous les motifs de répétitions ne se rencontrent cependant pas à la même fréquence dans les génomes. En effet, il y a une nette prédominance des répétitions de di-nucléotides, couvrant à elles-seules 0,5% du génome (*Consortium 2001*). De plus, selon le motif répété, le biais de répartition entre introns et exons varie. Par exemple, chez l'humain et plus généralement chez les primates, dans les régions intergéniques et les introns, les répétitions de mononucléotides sont deux fois plus fréquentes que celles de di- ou tétra-nucléotides (*Tóth et al. 2000*). Ou encore, les motifs de 3 et 6pb sont prédominants dans les exons. De manière générale, sur 1 mégabase, les tri- et pentanucléotides représentent entre 0,05-0,1% de la séquence, tandis que les di-,tetra- et hexanucléotides représentent entre 0,15-0,4% (*Subramanian et al. 2003*).

Les microsatellites ont longtemps été considérés comme de « l'ADN poubelle », entités neutres dans l'ADN. A présent, il a été démontré que les microsatellites sont primordiaux pour l'intégrité des génomes. Ils sont particulièrement importants dans le maintien de la structure de l'ADN, *via* la formation de structures secondaires que leur repliement impose *(Li et al. 2002)* : ces repliements influencent l'expression des gènes. Nous nous proposons dans la suite de ce chapitre de développer plus particulièrement le cas d'une famille de microsatellites particulière : les répétitions de trinucléotides.

1.2.2. LES REPETITIONS TRINUCLEOTIDES DANS LE GENOME HUMAIN

Plus de 2000 gènes ont au moins un exon associé avec au minimum 4 répétitions d'un motif trinucléotide. Ces répétitions sont notamment retrouvées dans les gènes codant des protéines nucléaires (*Mirkin 2007*). Les triplets répétés (TNRs) sont 2 fois plus fréquents dans les exons que dans les introns sauf sur le chromosome Y. A noter cependant que chez l'Homme, 4,6% des séquences 3'UTR et 31,1% des régions 5'UTR contiennent des triplets répétés (Subramanian *et al.* 2003). La répartition des triplets dans le génome est biaisée selon la nature des TNRs. Par exemple, les triplets les plus fréquemment rencontrés sur l'ensemble du génome humain sont les motifs AAA et TTT. Mais les motifs répétés les plus présents dans les exons correspondent à des CnG, ce qui laisse supposer une pression de sélection positive de ces répétitions (Kozlowski *et al.* 2010). Dans les régions utroniques et intergéniques, c'est le motif AAT qui est le plus fréquent (0,03%), quant aux régions UTRs, ce sont majoritairement des répétitions CCG et CGG, AAT et AAC qui y sont retrouvées. Les zones de répétitions CTG/CAG sont quant à elles plus rares, estimées à 900 dans tout le génome humain. *(Wren et al.* 2000, Subramanian *et al.* 2003).

Les microsatellites sont des zones présentant de hauts taux de mutations. En effet, leur taux de mutation est compris entre 10⁻³ et 10⁻⁴ là où celui estimé sur l'ensemble du génome humain est en moyenne de 10⁻⁸ par nucléotide (Cavalli-Sforza *et al.* 2006). Cependant, dans la population générale, l'on considère que les microsatellites sont stables. Certains motifs sont associés avec des maladies génétiques en cas d'instabilité anormale au cours de la vie des individus (instabilité somatique) ou d'une génération à l'autre (instabilité intergénérationnelle).

II. LES MALADIES HUMAINES A REPETITIONS DE NUCLEOTIDES

II.1. APERÇU GENERAL

Il existe une guarantaine de pathologies humaines neurodégénératives ou neurologiques liées à des séquences répétées (Pearson et al. 2005, Paulson 2018). Ces mutations apparaissent sur des régions de motifs répétés et polymorphes, mais généralement stables au sein de la population générale. Ces maladies se déclarent lorsqu'un nombre seuil de répétitions est dépassé, généralement à la suite d'une addition anormale de motifs répétés (expansion). Cette valeur seuil est spécifique de chaque maladie. Les répétitions impliquées sont des motifs de tailles différentes variant entre 3pb (trinucléotides CTG dans la Dystrophie Myotonique de type 1) et plusieurs kpb (unité D4Z4 de 3.3kpb dans la Dystrophie Musculaire Fascioscapulo-humérale). Les répétitions peuvent être de natures différentes (riches en A,T,G ou C) et peuvent être présentes sur des séquences codantes ou non codantes. Les conséquences de ces répétitions sont variables, elles peuvent causer un gain ou une perte de fonction de l'ARN ou de la protéine issus du gène muté (voir figure 2). Les maladies à répétitions de nucléotides sont divisées en 2 catégories. Les maladies à séquences répétées stables (exemple : séquences répétées de poly-alanine) et les maladies à séquences répétées dynamiques. Les maladies à répétitions dynamiques témoignent d'une instabilité du nombre de répétitions, avec généralement un biais vers les expansions, au cours de la vie des patients (instabilité somatique) et de génération en génération (instabilité intergénérationnelle).



Figure 2: Exemples de localisation génique et conséquences de différentes natures de répétitions de nucléotides.

(*) Les répétitions CTG chevauchent la région 3'UTR de DMPK et la région promotrice de SIX5. (**)Le gène sur lequel des répétitions CTG sont localisées possède 3 cadres de lectures différents. Selon le cadre utilisé et le sens de la transcription les répétions auront des conséquences différentes. EPM1 (Maladie de Unverricht Lundborg) ; SLA (Sclérose Latérale Amyothrophique) ; DM2 (Dystrophie Myotonique de type 2). Pour les autres abréviations, se référer aux tableaux 2,3,4. Cette thèse concernant l'instabilité des triplets CTG, les descriptions présentées cidessous seront orientées vers les maladies associées aux trinucléotides répétés. Les triplets répétés ont été liés à des pathologie génétique humaine à partir des années 1990, avec les répétitions CGG/CCG sur le gène *FMR1* dans le syndrome de l'X Fragile (Oberle *et al.* 1991, Verkerk *et al.* 1991) ; l'expansion CAG/CTG dans le récepteur à androgène, impliqué dans l'amyotrophie spino-bulbaire SBMA (Verkerk *et al.* 1991) ; et les répétitions CTG/CAG dans le gène *DMPK* impliquées dans la DM1 (Brook *et al.* 1992, Fu *et al.* 1992, Mahadevan *et al.* 1992). Dans la suite de ce mémoire, les pathologies seront organisées en 2 groupes, en fonction de la localisation des répétitions, à savoir sur une région codante ou non codante. Les tableaux récapitulatifs des pages suivantes présentent les principales caractéristiques de ces pathologies.

Pathologies	Phénotype principal	Gène	Localisation chromosomique	Position : taille de l'expansion	Protéine mutée	Fonction initiale de la protéine	Conséquence sur la protéine	Mode de transmission
Synpolydactilie type II (SPD)	Malformation des membres	HOXD13	2q31	Exon 1 : (15 → 22-29)	Homeobox D13	FT impliqué dans le développement des voies génito- urinaires et des membres	Gain de fonction	Autosomique dominant
Dysplasie cléridocranienne (CCD)	Dysplasie squelettique	RUNX2	6p21	Exon3 : $(17 \rightarrow 27)$	Runt related transcription factor 2	Facteur de transcription impliqué dans la différenciation des ostéoblastes et du développement squelettique	Perte de fonction	Autosomique dominant
Dystrophie musculaire occulopharyngée (OPMD)	Ptosis, dysphagie et faiblesse des membres	PABPN1	14q11	Exon 1 : (10 → 17) Exon 1 : 11	Poly(A)-binding protein nuclear 1	Protéine de liaison à la queue polyA des ARNm pour l'exportation nucléocytoplasmique, protéine impliquée dans la différenciation musculaire	Gain de fonction	Autosomique dominant Autosomique récessif
Holoprosencéphalie (HPE)	Développement anormal et malformation du système nerveux central (CNS)	ZIC2	13q32	Exon 3 : $(15 \rightarrow 25)$	Zinc finger protein 2	Facteur de transcription impliqué dans le développement du CNS	Perte de fonction	Autosomique dominant
Syndrome main-pied- génital (HFGS)	Anomalies squelettiques et malformations urogénitales	HOXA13	7q14	Exon 1 : Tract# 1 (14 \rightarrow 22) Tract# 2 (12 \rightarrow 18)	Homeobox A13	Facteur de transcription impliqué dans le développement des voies génito-urinaires et des membres	Perte ou gain de fonction ?	Autosomique dominant
Syndactylie blepharophimosis ptosis-epicanthus inversus (BPEIS)	Anomalies des paupières et insuffisance ovarienne prématurée	FOXL2	3q23	Exon unique : (14→19, 22 et 24)	Fork Head box L2	Facteur de transcription impliqué dans la fonction ovarienne et le développement des paupières	Perte de fonction partielle	Autosomique dominant
Retard mental lié à l'X avec déficit somatotrope	Nanisme et déficit en hormone de croissance	SOX3	Xq26	Exon unique : (15→22, 26)	Sex-determining Region of Y -related HMG-box 3	Facteur de transcription impliqué dans le Développement neural et la fonctionnalité de l'axe hypothalamo-hypophysaire	Perte de fonction	Récessif lié à l'X
Retard mental lié à l'X syndromique et non syndromique (XLMR)	Retard mental associé au syndrome de West ou à des mouvements dystoniques des mains et à une dysarthrie liée au syndrome de Partington	ARX	Xp22	Exon 2 : Tract# 1 (16 \rightarrow 23) Tract# 2 (12 \rightarrow 20)	Aristaless-related homeobox	Facteur de transcription impliqué dans le développement cérébral	Gain de fonction	Récessif lié à l'X
Syndrome d'hypoventilation centrale congénitale (syndrome d'Ondine ou syndrome de Haddad) (CCHS)	Anomalies du système nerveux autonome	PHOXB2	4p12	Exon 3: (20 → 25-33)	Paired-like homeobox 2b	Facteur de transcription impliqué dans la régulation du système nerveux autonome (ANS)	Perte de fonction	Autosomique dominant

Tableau 2: Aperçu des principales caractéristiques des pathologies dues à des répétitions GCG dans des régions codantes (maladies polyalanines). Adapté de Messaed et al. 2009.

Pathologies	Phénotype principal		Gène	Localisation chromosomique	Position : taille de l'expansion	Protéine mutée	Fonction initiale de la protéine	Conséquence sur la protéine	Mode de transmission
Atrophie musculaire spinale et bulbaire liée à l'X (SBMA, Maladie de Kennedy)	Atrophie musculaire, tren troubles end	nblements des mains, ocriniens	AR	Xq11-q12	Exon 1 : 38 → 70	Androgen receptor	Récepteur nucléaire aux stéroïdes	Gain de fonction	Récessif lié à l'X
Maladie de Huntington (HD)	Perte progressive du contrôle des mouvements, perte des fonctions cognitives, démence		HTT	4p16.3	Exon 1: $36 \rightarrow 121$	Huntingtin	Signalisation, transport	Gain de fonction	Autosomique dominant
Atrophie dentato-rubro- pallido-luysienne (DRPLA)	Myoclonus, épilepsie, ataxie cérébelleuse, démence, troubles psychiatriques		ATN1	12p13	Exon 5: 49 → 88	Atrophin-1	Kinase, Co- represseur de transcription	Gain de fonction	Autosomique dominant
Ataxie spinocérébelleuse de type 1 (SCA1)	Troubles comportementaux, dysphagie et amyotrophie	Toutes les SCAs présentent des troubles de coordination du mouvement, un nystagmus, une dysarthrie cérébelleuse et des troubles du langage	ATXN1	6p23	Exon : $39 \rightarrow 88$	Ataxin-1	Facteur de transcription	Gain de fonction	Autosomique dominant
SCA2	Mouvements choréiques, hyporéflexie, tremblements		ATXN2	12q24	Exon 1 : 32 → 77	Ataxin-2	Métabolisme des ARN	Gain de fonction	Autosomique dominant
SCA3	Tableaux cliniques hétérogènes, faiblesses musculaires, perte sensorielle, amyotrophie, ataxie, troubles apparentés à la maladie de Parkinson		ATXN3	14q24-q31	Exon 3 : 55 → 86	Ataxin-3	Dé-ubiquitinase, cystéine protéase	Gain de fonction	Autosomique dominant
SCA6	Atteintes diverses liées au cérébellum		CACNA1A	19p13	Exon 47 : 21 \rightarrow 33	Cav2.1	Sous-unité canal calcique	Gain de fonction	Autosomique dominant
SCA7	Ataxie, perte de la coordination des mouvements et de la marche, dégénérescence rétinienne		ATXN7	3p21-p12	Exon 3 : $38 \rightarrow 200$	Ataxin-7	Sous-unité complexe de transcription SAGA	Gain de fonction	Autosomique dominant
Huntington like disease 2 (HDL2, SCA17)	Démence, troubles apparentés à la maladie de Parkinson, mouvements involontaires, hyperreflexion		TBP	6p27	Exon 3 : 45 → 63	TATA- binding protein	Facteur de transcription	Gain de fonction	Autosomique dominant

Tableau 3: Caractéristiques des pathologies dues à une expansion CAG dans des régions non codantes (poly-glutamine, polyQ) adapté de Andrew P.Lieberman et al. 2019

Pathologies	Phénotype principal	Gène	Localisation chromosomique	Nature de la répétition	Position : taille de l'expansion	Conséquence sur ARN	Fonction initiale de la protéine	Conséquence sur la protéine	Mode de transmission
Syndrome de l'X fragile (FRAXA/FXS)	Symptômes apparaissant durant l'enfance, Retard dans l'apprentissage des fonctions motrice et/ou du langage, déficit intellectuel, dysmorphies, trouble du comportement	FMR1	Xq27.3	CGG	>200	Perte de fonction	Régulateur de la transcription	Diminution de la protéine	Dominant lié à l'X
Syndrome de l'X fragile avec tremblement et ataxie (FXTAS)	Symptômes apparaissant chez l'adulte, tremblement, ataxie cérébelleuse progressive de la marche, manifestations psychiatriques	FMR1	Xq27.3	CGG	55-200	Gain de fonction	Régulateur de la transcription	Diminution de la protéine	Dominant lié à I'X
Syndrome de l'insuffisance ovarienne liée à l'X fragile (FXPOI)	Taux constamment élevés de FSH entrainant une dérégulation de la fonction ovarienne : cycles irréguliers, infertilité, ménopause précoce	FMR1	Xq27.3	CGG	55-200	Gain de fonction	Régulateur de la transcription	Diminution de la protéine	Dominant lié à I'X
Syndrome de l'X fragile de type E (FRAXE)	Troubles intellectuels et cognitifs, troubles de l'apprentissage	FMR2	Xq28	CCG	>200	Perte de fonction	Régulateur de la transcription	Diminution de la protéine	Dominant lié à I'X
Ataxie de Friedreich (FRDA)	Symptômes apparaissant dans l'enfance, ataxie des membres supérieurs, dysarthrie, nystagmus, trouble de l'audition, dysphagie, perte de la sensibilité distale	Frataxine	9q13-21.1	GAA	Intron : 200 → 1700	Perte de fonction	Facteur d'assemblage de protéines à noyau fer/cuivre	Diminution de la protéine	Autosomique Récessif
SCA12	Tremblements d'action, ataxie cérébelleuse, démence	PPP2R2B	5q31-33	CAG	Promoteur : 6 → 78 variables selon ethnie	Perte ou gain de fonction	Phosphatase	Diminution de la protéine	Autosomique Dominant
SCA8	Ataxie cérébelleuse, troubles cognitifs et psychiatriques		13q21	CTG	3'UTR : >74	Gain de fonction	Régulateur de la transcription	Diminution de la protéine	Autosomique Dominant
Maladie de Huntington de type 2 (HDL2)	Troubles psychiatriques, Mouvements choréiques, dystonie	Junctophiline 3	16q24.3	CTG	Exon épissé : 66 →78	Gain de fonction	Facteur d'assemblage de complexes membranaire et du Réticulum Endoplasmique	Diminution de la protéine et gain de fonction	Autosomique Dominant
Dystrophie myotonique de type 1 (DM1)	Myotonie, dégénérescence neuromusculaire	DMPK	19q13.3	CTG	3' UTR : 50 → 2000	Gain de fonction	Sérine-thréonine-kinase	Diminution de la protéine	Autosomique Dominant

Tableau 4: Caractéristiques des pathologies dues à une expansion de triplets dans une région non codante.

II.2. PATHOLOGIES CAUSEES PAR DES TRIPLETS REPETES DANS LES EXONS

Les répétitions anormales situées dans les régions codantes sont transcrites et traduites. Ces mutations pourront ainsi avoir des conséquences au niveau de l'ARN et de la protéine produits par le gène muté.

II.2.1. MALADIES A EXPANSION D'ALANINE (GCG)

Les expansions alanines sont codées par des répétitions GCG. La Synpolydactylie de type II a été la première maladie humaine associée à des expansions anormales d'alanine (Muragaki *et al.* 1996, Goodman *et al.* 1997). À ce jour, 9 maladies humaines associées aux répétitions GCG sont identifiées et sont connues pour généralement provoquer une dégénérescence neuromusculaire sévère et des dysfonctionnements neuronaux : leurs principales caractéristiques sont résumées dans le tableau 2. Dans la population générale, le nombre de répétitions GCG est peu polymorphe et les répétitions sont souvent interrompues par des triplets GCN (N=A, T ou C) et n'excédent jamais 20 répétitions. (Warren *et al.* 1997, Messaed *et al.* 2009). Ces répétitions deviennent pathologiques une fois un nombre seuil de répétitions atteint (nombre spécifique à chaque maladie, voir tableau 2), avec une corrélation positive entre la taille des répétitions et la sévérité des symptômes. (Amiel *et al.* 2004, Albrecht *et al.* 2005, Carroll *et al.* 2017) Les répétitions alanines sont très fréquemment observées dans des facteurs de transcription et entraînent un gain ou une perte de fonction de protéines selon les pathologies (voir tableau 2).

Les expansions alanines mènent à l'agrégation des protéines mutantes soit dans le cytoplasme, soit dans le noyau, soit dans les 2 compartiments en même temps. La formation d'agrégats est positivement corrélée à la longueur des répétitions (Albrecht et al. 2004, Caburet et al. 2004, Utsch et al. 2007). Les répétitions alanine mènent soit à un gain soit à une perte de fonction. Afin d'expliquer ces différentes conséquences, plusieurs mécanismes ont été proposés. Tous se basent sur le repliement et l'agrégation anormaux des protéines mutantes. Des expériences dans un modèle cellulaire issu de patients atteints de dystrophie musculaire occulopharyngée (OPMD) ont démontré que des protéines chaperonnes (HSP40 et HSP70) sont séquestrées par les protéines PAPNB1 mutantes agrégées dans le cytoplasme (Abu-Baker et al. 2003). Cette co-localisation anormalement stable avec des protéines chaperonnes peut être représentative de différents mécanismes : (1) les protéines chaperonnes arrivent à replier correctement les protéines et celles-ci peuvent alors être fonctionnelles, (2) les protéines chaperonnes échouent à replier correctement les protéines et les orientent vers le protéasome pour y être dégradées, (3) les protéines mutantes échappent aux mécanismes en aval de la fixation des protéines chaperonnes ainsi qu'au protéasome et vont former des agrégats qui confèrent un gain de fonction aux protéines. En effet, il a été

démontré dans les Syndrome- Main-Pied-Génital (Utsch et al. 2007) et de Synpolydactylie (Albrecht et al. 2004) que les protéines mutantes deviennent capables de séquestrer leur version normale (HOXD13 ou HOXA13 respectivement). Ces observations expliquent la « perte de fonction » de la protéine normale par « un gain de fonction » de la protéine mutante. Récemment, il a été démontré par des expériences de dichroïsme circulaire et de microscopie électronique, que les expansions d'alanine imposent un repliement des protéines sous forme d'hélice- α . La configuration en hélice- α rend possible l'agrégation d'un nombre variable de protéines entre-elles, formant des clusters à configuration et stabilité également



Figure 3: Observation et schématisation d'agrégats.



variables (voir figure 3). Selon la configuration de l'agrégat, les protéines optent pour un repliement anormal (voir figure 3) qui va entrainer la perte d'interaction avec un ligand (perte de fonction) ou au contraire créer de nouvelles interactions (gain de fonction) (Polling *et al.* 2019). Il est probable que les pertes et gains de fonction des protéines décrits ci-dessus soient des mécanismes complémentaires entre eux participant tous ensembles à la physiopathologie des maladies à expansion d'alanine.

Contrairement aux autres maladies à triplets, les répétitions GCG restent relativement stables au cours de la vie des patients et lors des transmissions (Pearson *et al.* 2005). L'instabilité des répétitions de trinucléotides dépend effectivement en partie de leur nature et de leur longueur (voir chapitre III.3). Or, la plus longue répétition de poly-alanine décrite à ce jour correspond à une répétition de 33 alanines dans le cadre du Syndrome d'hypoventilation centrale congénitale (Matera *et al.* 2004). À titre de comparaison, cela correspond à une taille de répétition qui serait normales pour les répétitions poly-glutamines associées à des pathologies par exemple, et c'est un nombre de répétitions très faible comparées aux plus de 1000 répétitions observées dans le cas de répétitions de triplets CTG dans la Dystrophie Myotonique de type 1 (DM1). La présence d'interruptions régulières est également un facteur pouvant expliquer la stabilité de ces répétitions (voir chapitre III.3.).

L'observation de recombinaisons homologues inégales dans des cas de Dystrophie musculaire occulo-pharyngée (OPMD), de Synpolydactilie de type II (SPD) et du Syndrome main-pied-génital laissent supposer que ce mécanisme de recombinaison, lorsqu'il implique deux allèles normaux, est la source d'expansions anormales de poly-alanine (Warren 1997, Nakamoto *et al.* 2002, Utsch *et al.* 2002, Robinson *et al.* 2005). Cependant, il a été observé des cas de Syndrome d'hypoventilation centrale congénitale où la taille des expansions d'alanines ne pouvait pas être expliquée par une recombinaison homologue inégale. En effet, les tailles des allèles parentaux normaux n'étaient pas compatible avec l'expansion observée chez le patient (Trochet *et al.* 2007). De plus, les auteurs ont révélé une mosaïque somatique chez des patients asymptomatiques et pourtant porteurs d'un allèle mutant. Ces observations laissent supposer l'existence de mécanismes d'instabilités communs entre les maladies à expansions de poly-alanine et les maladies à répétitions instables (décrites ci-après).

Il existe une vingtaine de maladies humaines associées à des « mutations dynamiques » de triplets (les mécanismes d'instabilité des triplets répétés serons traités dans le chapitre III.). Les motifs répétés peuvent être de différentes natures et le nombre seuil de répétitions pathologique dépend de chacune des maladies. Une fois ce seuil atteint, les répétitions deviennent instables, de génération en génération (instabilité intergénérationnelle) et au cours de la vie des patients dans les tissus (instabilité somatique). Les répétitions anormales peuvent survenir dans les régions 5' UTR, les exons, les introns, et les régions 3' UTR (voir figure 2). Les conséquences de ces répétitions anormales vont être différentes selon leur nature, le nombre de répétitions, et leur localisation sur les gènes.

Toutes les maladies à triplets instables dans les régions exoniques sont dues à des répétitions CAG, aboutissant à un gain de fonction des protéines mutantes qui sont allongées avec une chaîne de poly-glutamine.

II.2.2. MALADIES A EXPANSION DE GLUTAMINES (POLY-Q)

À ce jour, 9 maladies humaines sont associées aux répétitions anormales de glutamine. Toutes sont autosomiques dominantes (à l'exception du SBMA qui est lié au chromosome X, voir tableau 3). Les maladies Poly-Q sont neurodégénératives, menant à une perte des fonctions motrices, et cognitives (voir (Pringsheim *et al.* 2012) pour revue). Les répétitions CAG sont instables une fois une valeur seuil pathologique atteinte, cette valeur seuil étant spécifique à chaque maladie (généralement une quarantaine de CAG, voir tableau 3). Ces maladies présentent une instabilité intergénérationnelle et somatique biaisées vers les expansions. Généralement, l'âge d'apparition des symptômes est inversement corrélé au nombre de répétitions CAG hérité : plus les répétitions sont nombreuses, plus les symptômes sont précoces et sévères. Des cas particulièrement sévères se sont déjà manifestés sur des

patients de moins de 20 ans, notamment pour la maladie de Huntington, dans le cas d'expansion CAG exceptionnellement longues à la suite d'une transmission paternelle (Van Dijk *et al.* 1986, Telenius *et al.* 1993). Ces maladies présentent un phénomène d'anticipation (aggravation et âge de plus en plus précoce des symptômes au cours des générations), bien que celui-ci reste modéré lorsqu'il est comparé à celui observé dans les autres maladies à mutations dynamiques comme la DM1. Le plus fort phénomène d'anticipation dans les maladies à poly-glutamine est observé dans la DRPLA, maladies quasi-exclusivement retrouvée au Japon (effet fondateur), où l'instabilité intergénérationnelle est biaisée dans 80% des cas vers les expansions, avec en moyenne +13 CAG par transmission, avec une instabilité et un gain de (n)CAG plus importants lors de transmissions paternelles (Ikeuchi *et al.* 1995). Cependant, une étude sur 188 patients porteurs de différentes SCAs nuance ces propos : en effet, aucune corrélation significative entre l'âge d'apparition des symptômes et le nombre de répétitions CAG hérité n'y a été trouvée (Lone *et al.* 2014). Généralement, les symptômes apparaissent à l'âge adulte et suivent une lente progression.

Les expansions poly-glutamines mènent à l'agrégation de protéines mutantes suite à un mauvais repliement de celles-ci, et cela malgré le recrutement de protéines chaperonnes (Ciechanover *et al.* 2003). Les protéines poly-Q s'agrègent sous forme d'inclusions nucléaires notamment dans les neurones (SCA1, 7 et 17), ou forment des agrégats dans le cytoplasme (SCA2, SCA6) voire dans les 2 compartiments en même temps (SCA3, HD, SBMA, DRPLA) (Gatchel *et al.* 2005). Ces agrégats peuvent être de différentes tailles, aboutissant à de longues structures fibrillaires aberrantes (voir figure 3). Les conséquences physiopathologiques de ces agrégats sont encore aujourd'hui sujet à débat.

La première hypothèse serait que les inclusions de protéines avec expansions poly-Q soient vecteurs de mort cellulaire. De nombreuses études, principalement dans le cadre de la maladie de Huntington, démontrent un aspect délétère des agrégats de protéines mutantes sur de nombreuses fonctions cellulaires (voir (Lieberman *et al.* 2019) pour revue). Par exemple, les agrégats déstabilisent la matrice et l'enveloppe nucléaires entrainant ainsi la mort cellulaire (Skinner *et al.* 1997, Liu *et al.* 2014). Les agrégats séquestrent des facteurs capitaux pour la survie cellulaire en co-séggrégeant avec les ribo-protéines (Ramdzan *et al.* 2017), en délocalisant des facteurs de transcription comme TBP (Boutell *et al.* 1999) et des protéines du transport nucléocytoplasmique (Woerner *et al.* 2016) ou encore des protéines du transport axonal (Orr *et al.* 2007). Il a aussi été démontré que ces agrégats dérégulent l'adressage de protéines vers le Réticulum Endoplasmique, déstabilisant ainsi ce dernier (Bäuerlein *et al.* 2017). De plus, les agrégats en grand nombre résistent au système de dégradation par le protéasome, perturbant ainsi cette voie métabolique à l'échelle de la cellule entière (Ciechanover *et al.* 2003). Enfin, il a été démontré que les inclusions nucléaires d'huntingtine

mutante fixent et transforment la structure chromatinienne, dérégulant les interactions de l'ADN avec les protéines nucléaires et la bonne répartition chromosomique au cours de la neurogénèse (Li *et al.* 2016, Ruzo *et al.* 2018).

Cependant, malgré toutes ces démonstrations, le doute demeure sur l'aspect uniquement pathologique des agrégats de protéines à expansion de glutamine. En effet, des travaux sur des échantillons de cerveaux post-mortem et des modèles murins ont démontré qu'il n'y a pas de bonne corrélation entre l'atteinte tissulaire et le nombre d'inclusions d'huntingtine mutante (Saudou et al. 1998, Kuemmerle et al. 1999). De plus, dans un modèle cellulaire de la maladie de Huntington, la survie des neurones en culture est améliorée en présence d'inclusions nucléaires (Arrasate et al. 2004). Il a par ailleurs été prouvé dans des modèles de souris de la SCA1 et SCA7, qu'en absence d'inclusion nucléaire de protéines allongées, la maladie progresse plus rapidement (Cummings et al. 1999, Watase et al. 2002, Yoo et al. 2003). Enfin, les protéines huntingtine mutantes solubles sont connues pour déclencher l'apoptose rapidement, tandis que les agrégats inhibent l'apoptose et séquestrent des éléments toxiques pour la cellule (comme des structures prions), mais déclenchent cependant plus tardivement la nécrose (Orr 2001, Riley et al. 2004, Ramdzan et al. 2017). L'ensemble de ces travaux tendent à conclure que les protéines mutantes solubles sont toxiques, tandis que les agrégats pourraient être les témoins de mécanismes protecteurs de la cellule.

Il est très fortement probable que les agrégats soient à la fois pathologiques et protecteurs. Les agrégats seraient tout d'abord le résultat de mécanismes protégeant la cellule des protéines allongées solubles et d'autres molécules toxiques. Les agrégats seraient bénéfiques tant que leur nombre et leur taille restent « gérables » pour la cellule (notamment via le système de destruction par le protéasome). Puis, en s'accumulant, suite à l'accumulation de répétitions CAG au cours de la vie des patients, les agrégats deviennent ingérables et résistant aux mécanismes de destruction. Ils s'accumulent alors dans les cellules, jusqu'à devenir toxiques de diverses manières (voir paragraphes précédents) puis déclenchent la nécrose (Tallaksen-Greene *et al.* 2003, Ramdzan *et al.* 2017).

Les maladies à expansions d'alanine et de glutamine ont des caractéristiques qui leurs sont propres (nature des répétitions, seuils pathologiques...). Pour autant, elles partagent des mécanismes de toxicité communs, basés sur le mauvais repliement des protéines et la mauvaise prise en charge de celle-ci par la cellule. *A contrario*, les maladies à mutation dynamiques qui seront décrites ci-après sont dues à des mutations dans des régions non codantes : leur pathogénicité sera uniquement liée à l'ADN ou aux ARN mutants générés par les triplets répétés.

II.3. PATHOLOGIES CAUSEES PAR DES TRIPLETS REPETES DANS DES REGIONS NON CODANTES DES GENES

Il a été décrit une vingtaine de pathologies humaines dues à des mutations dynamiques dans les régions non codantes des gènes. La majorité d'entre-elles sont causées par des triplets répétés. Les principales caractéristiques de ces pathologies sont répertoriées dans le tableau 4. Dans les paragraphes suivants, nous présenterons un aperçu des maladies à triplets répétés instables en fonction de la nature des répétitions impliquées, avant de recentrer le discours sur la Dystrophie Myotonique de type 1, sujet de ce mémoire.

II.3.1 EXPANSIONS CGG/CCG : SYNDROME DE L'X FRAGILE ET MALADIES ASSOCIEES

Avec une fréquence de 1/5 000 pour les hommes et jusqu'à 1/8 000 pour les femmes, le syndrome de l'X fragile et les syndromes qui y sont associés sont les causes héréditaires de déficience intellectuelle les plus répandues dans le monde, ainsi que les maladies monogéniques transmettant le plus fréquemment des troubles de la sphère autistique. Cependant, l'on distingue des effets fondateurs très marqués pour ces maladie. En Colombie, précisément dans la ville de Ricaurte, 1 homme sur 25 et 1 femme sur 49 sont porteurs de l'allèles avec une mutation dite « complète » (voir paragraphe suivant). Par opposition, on ne relève en Irlande qu'une fréquence de 1/ 10,619 chez les hommes et de 1/43,540 chez les femmes pour cet allèle. Les enfants naissant avec une mutation complète sont généralement hypotoniques, et présentent notamment des troubles du développement et un retard dans l'apprentissage du langage (voir (Hagerman *et al.* 2017, Rajaratnam *et al.* 2017) pour revues).

Le **Syndrome de l'X Fragile (FXS)** est dû à la mutation « complète » dans la région 5'UTR du gène *FMR1* porté sur le chromosome X et codant pour la protéine FMRP. La mutation est dite « complète » lorsque *FMR1* est porteur d'au moins 200 répétitions CGG (contre 6 à 52 dans la population générale). Les allèles pré-mutés (55 à 200 CGG) peuvent transmettent d'autres maladies génétiques, cliniquement distinctes : les hommes développent le plus souvent un **Syndrome de l'X Fragile avec Tremblement et Ataxie** (FXTAS) tandis que les femmes sont le plus souvent atteintes du **Syndrome de l'Insuffisance Ovarienne Précoce liée au X fragile** (FXPOI). Cependant, la pénétrance de ces syndromes associés à l'X fragile est incomplète : tous les porteurs de pré-mutation ne sont pas symptomatiques. À partir de 55 répétitions, les triplets CGG sont instables, augmentant d'une génération à l'autre avec notamment une influence du sexe du parent transmetteur et de la longueur de la taille de répétition héritée. En effet, les mutations « complète » sont exclusivement transmises par les mères. (Fu *et al.* 1991, Moutou *et al.* 1997, Tassone *et al.* 2013, Yrigollen *et al.* 2014).

La mutation complète entraine une sur-méthylation locale de la région promotrice du gène *FMR1, ce* qui diminue la transcription de ce gène. Il en résulte un déficit pathologique

(perte de fonction) en FMRP, protéine régulant initialement la transcription de facteurs impliqués dans le maintien des synapses. Le mécanisme derrière cette dérégulation n'est pas bien connu. Mais une étude dans des cellules souches d'embryons humains a mis en évidence que la désactivation du gène *FMR1* commençait à la 11^{ème} de semaine de gestation. Dans cette étude, l'inactivation du gène est liée au recrutement de l'histone H3 suite à l'accumulation anormale d'hybrides ADN/ARN au niveau des répétitions CGG (Colak *et al.* 2014).

Contrairement au FXS, la pré-mutation entraine une élévation du niveau de transcription de *FMR1*, un gain de fonction toxique de l'ARN et une diminution paradoxale de la production de protéine FMRP (Primerano et al. 2002). Une étude sur des cerveaux postmortem de patients a mis en évidence la présence d'inclusions nucléaires d'ARN pré-mutés dans les cellules neuronales. Ces agrégats contiennent les ARNm FMRP pré-mutés ainsi que des protéines chaperonnes, des ribo-protéines, MBNL1 et des protéines filamentaires (Tassone et al. 2004, Iwahashi et al. 2005). Contrairement au FXS à mutation complète, les symptômes résultant d'une mutation incomplète apparaissent à l'âge adulte (environs 40 ans). À noter, qu'il a récemment été décrits 2 cas de patients porteurs de pré-mutation (65 et 66 CGG), présentant des troubles neurologiques mais pas de tableau clinique correspondant au FXTAS et, surtout, ils ne présentaient aucune inclusion d'ARN mutants. Une explication aurait pu être le faible nombre de répétitions CGG chez ces patients, mais, dans la même étude, les auteurs ont réussi à détecter des inclusions nucléaires chez des patients avec 63 et 67 répétitions. Ces cas laissent suggérer que les inclusions nucléaires ne sont pas les seuls éléments responsables de la pathogénèse causées par les pré-mutations sur FMR1 (Martinez-Cerdeno et al. 2017).

Par ailleurs, il a été révélé des cas où les patients ont une mosaïque de la taille des allèles mutants (Pretto *et al.* 2014), c'est-à-dire que, dans les cellules d'un même patient, il a été retrouvé soit des allèles pré-mutés ou des allèles à mutation « complète ». D'autres patients, porteurs exclusivement de la mutation « complète » présentent une mosaïque dans le statut de méthylation (Pretto *et al.* 2014) de la région promotrice de *FMR1* : toutes les cellules 'un même individu ne présentent pas le même profil de méthylation de cette région. Les patients porteurs de ces mosaïques ont eux aussi une faible production de FMRP et présentent un tableau clinique moins sévère que les patients porteurs uniquement de la mutation complète. Ces mosaïques troublent le diagnostic et les dénominations des syndromes liés à l'X Fragile. En effet, une étude post-mortem sur différents tissus d'un patient très légèrement atteint (faibles troubles de l'attention et de l'apprentissage) et pourtant porteur d'une mutation « complète », présentait une mosaïque de taille <u>et</u> de méthylation (Taylor *et al.* 1999). Ces mosaïques sont d'importants modulateurs de la maladie et tendent à être considéré comme de réels biomarqueurs pour le pronostic de FXS, FXTAS et FXPOI.

Il existe aussi le **FRAXE** (Syndrome de l'X Fragile de type E) présentant un retard mental modéré comparé aux syndromes décrits plus hauts. Les symptômes sont dus à l'expansion de triplets **CCG** dans la région 5' UTR du gène *FMR2* : le nombre de répétitions chez les patients est >200 contre 4 à 42 CCG dans la population générale (Gecz *et al.* 1996). Ces répétitions aboutissent à la troncature ou l'absence de FMR2, et à l'hyper-méthylation de la zone promotrice de *FMR2* : l'expression de *FMR2* s'en retrouve diminuée (Gécz *et al.* 1997). De plus, l'expansion de CCG hyper-méthylée affecte l'expression du gène *FMR3*. En effet, *FMR3* est proche du gène *FMR2*, chacun de ces gènes étant transcrit dans un sens différent (transcription divergente): la méthylation des répétitions CCG en 5' UTR de FMR2 est donc située dans la région promotrice du gène *FMR3*, ce qui diminue la transcription de ce dernier (Gécz 2000). FMR2 est un facteur de transcription impliqué dans le développement du système nerveux et la plasticité synaptique. L'expansion de CCG sur *FMR2* est responsable des troubles autistiques présentés par les patients, le rôle de *FMR3* dans la pathologie reste cependant méconnu (Gécz 2000).

II.3.2. MALADIE A EXPANSIONS GAA : L'ATAXIE DE FRIEDREICH (FRDRA OU FA)

L'Ataxie de Friedreich est une maladie autosomique récessive due à une expansion GAA dans l'intron 1 du gène *frataxine (FRDA/FXN)* supérieure à 200 GAA contre 6-35 dans la population générale. Les allèles avec une répétition de GAA comprise entre 35 et 200 sont dits « prémutés ». L'expansion de GAA est généralement homozygote, mais il a aussi été identifié que dans de très rares cas, des mutations ponctuelles dans l'un des allèles de *FRDA/FXN*, associées à des expansions GAA hétérozygote sont coresponsables de la maladie (Orr *et al.* 2007, Willis *et al.* 2008, Sacca *et al.* 2011, Clark *et al.* 2019). L'ataxie de Friedreich est la forme d'ataxie héréditaire la plus fréquente : elle est responsable de la moitié des ataxies héréditaires, et de 75% de celles se manifestant avant 25 ans. Dans les populations occidentales, la prévalence de la FRDA varie entre 1/20 000 et 1/725 000 (voir (Bürk 2017) pour revue). La comparaison de données épidémiologiques et de bases de données gérées par des associations de patients ont permis de démontrer que les prévalences de FRDA les plus élevées se retrouvent en Irlande, France et Espagne du nord (Vankan 2013).

Les répétitions GAA entrainent l'extinction du gène *FRDA/FXN, aboutissant* à un déficit de la protéine frataxine. Les répétitions entrainent une dérégulation de l'expression de la frataxine en stabilisant la formation d'R-loop (hybrides ADN/ARN), qui provoquent des changements de marques épigénétiques et l'hétérochromatisation du gène FRDA/FXN (Groh *et al.* 2014, Groh *et al.* 2014). Normalement, la frataxine active l'assemblage de protéines à noyaux fer-soufre pour composer des complexes protéiques dans tous les compartiments cellulaires. Le déficit
en frataxine entraine surtout une diminution de ces complexes au niveau de la chaîne respiratoire mitochondriale. Ainsi, l'on observe chez les patients un dysfonctionnement mitochondrial progressif, menant à une accumulation de fer sur les membranes mitochondriales, à une plus grande fragilité des cellules face au stress oxydatif, jusqu'à causer la mort cellulaire, principalement dans les neurones sensoriels et le cervelet (Saveliev *et al.* 2003, Groh *et al.* 2014). Les conséquences sur le plan clinique, se manifestent notamment par une neuro-dégénérescence progressive menant à une spasticité et à l'absence de réflexe des membres inférieurs, une cardiomyopathie, un diabète sucré et de nombreuses difformités squelettiques (Pandolfo 2009).

La sévérité des symptômes est corrélée à la taille de la répétition de GAA héritée la plus courte. Autrement dit, plus la taille de la plus courte expansion GAA pathologique héritée est longue, plus les symptômes sont sévères et précoces (Dürr *et al.* 1996, Mateo *et al.* 2004). Il a par exemple été estimé qu'à chaque pallier de 100 GAA ajoutés, la maladie se déclare 2-3 ans plus tôt (Reetz *et al.* 2015). La répétition de GAA devient instable au-delà de 200 répétitions, et peut atteindre jusqu'à 1700 répétitions. Les expansions pathologiques les plus communes varient entre 600-900 GAA.

L'instabilité intergénérationnelle des triplets GAA pathogènes est particulièrement importante : jusqu'à +300 répétitions de rajoutés en 1 transmission. Cette instabilité dépend du sexe et de la taille l'expansion de GAA du parent transmetteur : les transmissions paternelles sont toujours biaisées vers les contractions de GAA tout comme les transmissions maternelles lorsque la mère porte plus de 800 GAA répétés. Par contre, lorsque la mère transmettrice présente moins de 800 GAA, l'instabilité intergénérationnelle est biaisée vers les expansions (De Michele *et al.* 1998). Différentes études sur des tissus de cohortes de patients ont démontré une instabilité somatique qui s'accentue au cours du temps. Ces analyses ont mis en évidence que l'instabilité somatique est tissu-spécifique, avec les plus grandes fréquences des plus grandes tailles d'expansion dans cœur, le pancréas et les ganglions de l'épine dorsale (De Biase *et al.* 2007, Long *et al.* 2017), tissus particulièrement atteints dans la FRDA.

II.3.3 MALADIE A EXPANSION CAG/CTG – SCA12

L'ataxie spino-cérébelleuse de type 12 (SCA12) est une maladie neurodégénérative, transmise sur le mode autosomique dominant. La SCA12 est causée par une expansion de triplet CAG dans la région 5'UTR du gène *PPP2R2B*, codant pour une sous-unité d'une sérine/thréonine phosphatase (PP2A) exprimée dans le cerveau (Holmes *et al.* 1999).

Le peu de données répertoriées à ce jour rend difficile la confirmation ou infirmation d'une corrélation entre les symptômes et la taille de triplets répétés (Merrill et al. 2012). Cela semble mettre en avant que la taille de la répétition CAG pathologique ne peut pas expliquer à elle seule les symptômes. De plus, il a été démontré chez une famille de patients japonais que le tableau clinique de la SCA12 pouvait se manifester sans qu'il n'y ait de répétitions CAG aberrantes (16-17 selon les patients). Ce cas laisse également supposer que la SCA12 n'est peut-être pas seulement dépendante des répétitions CAG. Par « mapping », les auteurs ont proposé 44 gènes candidats pouvant être impliqué dans la pathogénèse de la SCA12 dans chez ces patients. Parmi eux, HTR4 et ADRB2, deux gènes très fortement exprimés dans le cerveau mais aucune démonstration directe de leur implication n'a été faite (Sato et al. 2010). Par ailleurs, la très faible instabilité des répétitions impliquées dans la SCA12 (+/- 1 à 2 CAG par génération) n'est pas corrélée à la gravité et l'âge d'apparition des symptômes (Holmes et al. 1999, Fujigasaki et al. 2001, O'Hearn et al. 2001, Srivastava et al. 2001). Par exemple, dans l'étude de Lone et al., les plus petites tailles de répétitions (soient 6 CAG retrouvés chez 2 individus atteints par la SCA12) ne se retrouvent que chez les patients et non pas dans la population générale (Merrill et al. 2012).

À noter, en revanche, que la taille des répétitions CAG est positivement corrélée à l'expression de *PPP2R2B* (Chen *et al.* 2009, Lin *et al.* 2010, Kimura *et al.* 2011). Ceci peut s'expliquer par la proximité des répétitions CAG avec de nombreux sites d'initiation de la transcription de *PPP2R2B* : La présence de CAG au sein de ces séquences va moduler l'expression du gène *PPP2R2B*. En fonction du site utilisé pour la transcription du gène *PPP2R2B*, les conséquences de l'implication des répétitions sera différente. Les répétitions pourront être transcrites mais non traduites (entrainant un gain de fonction de l'ARN mutant), ou alors ne seront pas transcription par rapport à un autre est variable d'un patient à un autre, cette sélection est donc un élément pouvant expliquer la haute variabilité symptomatique de la SCA12.

II.3.4 MALADIES A EXPANSIONS CTG

Il existe plusieurs maladies à mutations dynamiques due à une expansion de triplets CTG dans différents gènes. Après de brefs aperçus de l'Ataxie Spinocérébelleuse de type 8 (SCA8) et la maladie de Huntington de type 2 (HDL2), ce chapitre présentera plus en détails la Dystrophie Myotonique de type 1 (DM1).

II.3.4.1 L'ATAXIE SPINOCEREBELLEUSE DE TYPE 8 (SCA8)

La SCA8 est une maladie neurodégénérative autosomique dominante, causée par une expansion de triplets CTG/CAG, adjacentes à des répétitions CTA/TAG, dans la région 3'UTR du gène ATXN8OS (anciennement connu sous le nom de SCA8). La région 3'UTR d'ATXN8OS se superpose à la région 5'UTR du gène ATXN8 transcrit dans le sens de inverse de ATXN8OS (Koob et al. 1999). Dans la population générale, les répétitions CTG sur le locus SCA8 sont polymorphes, comprises entre 15 et 50 répétitions (CTG/CAG)n combinées à des répétitions (CTA/TAG)n polymorphe stables. Chez les patients SCA8, le nombre de triplets répétés est généralement compris entre 50-250 répétitions. Il arrive que dans de rares cas, des patients présentent jusqu'à 1300 répétitions. A noter que seules les répétitions CTG/CAG sont instables. Cependant, la SCA8 a une pénétrance très réduite pour les allèles possédant entre 50 et 80 répétitions. Plusieurs facteurs peuvent influencer la pénétrance de la maladie dont le nombre de répétitions CTA/TAG, l'haplotype ou encore la présence d'interruptions (Ranum et al. 1999, Ikeda et al. 2004, Martins et al. 2005, Ikeda et al. 2008, Hu et al. 2017). Les répétitions CTG/CAG mutantes sont instables de génération en génération. Lors de transmissions maternelles, on observe un biais vers les expansions, lors des transmissions paternelles, l'instabilité intergénérationnelle est biaisée vers les contractions (Moseley et al. 2000). Les triplets CTG/CAG sur le locus SCA8 présentent également une instabilité somatique, que ce soit en cas de répétitions pathologiques ou normales. En effet une étude sur 20 familles asymptomatiques a démontré que les allèles porteurs de 23 à 28 répétitions (CTG/CAG et CTA/TAG) combinées étaient instables, avec de faibles contractions (1 CTG délété) ou présentaient une substitution en mosaïque d'un CTA par un CTG (Moseley et al. 2000, Silveira et al. 2000, Martins et al. 2005). Les répétitions sur le locus SCA8 seraient donc naturellement « instables », indépendamment de la longueur répétitions CTG/CAG. Cette observation soulignerait ainsi l'impact prédominant d'autres facteurs que la longueur des répétitions sur l'instabilité des triplets impliqués dans la SCA8.

Les transcrits ATXN8OS et ATXN8 sont exclusivement produits dans le cerveau (Janzen *et al.* 1999). Les transcrits **ATXN8OS** contiennent des répétitions CUG, qui vont agréger les ARNm mutants en formant des *foci* nucléaires. Les ARN mutants ont un gain de fonction qui leur permet de séquestrer MBNL1 (protéine impliquée dans l'épissage alternatif),

dérégulant ainsi l'épissage d'autres ARNm, comme par exemple l'épissage de transcrits codant une sous-unité du transporteur GABA-A (Daughters *et al.* 2009). Les transcrits *ATXN8* portent les répétitions CAG et codent pour une protéine mutante à expansion de poly-glutamine à traduction indépendante du mécanisme RAN. Ces protéines s'agrègent dans le noyau des cellules du cerveau. (Janzen *et al.* 1999, Moseley *et al.* 2006, Ikeda *et al.* 2008, Zu *et al.* 2011). Le rôle exact de ces ARNm et protéines mutantes ne sont pas décrits. Cependant, l'agglutination d'ARN sous forme de *foci* et de protéines sous forme d'agrégats sont connues pour être impliquées dans les symptômes d'autres pathologies à triplets répétés (DM1, HD...). Il est donc très probable que les transcrits *ATXN8OS* et *ATXN8* à expansion de CTG/CAG soient impliqués dans la pathogénèse de la SCA8.

II.3.4.2. LA MALADIE DE HUNTINGTON DE TYPE 2

La maladie de Huntington de type 2 (HDL2) est due à une expansion de CTG dans l'exon 2A épissé du gène *JPH3*. La HDL2 se transmet selon le mode autosomique dominant et n'est cliniquement pas distinguable de la maladie de Huntington. *JPH3* code pour la junctophiline 3 principalement exprimée dans le cerveau. Cette protéine participe aux jonctions entre la membrane cytoplasmique et la membrane du réticulum endoplasmique. Le gène *JPH3* présente de nombreux sites d'épissage. En fonction de l'épissage alternatif, les répétitions CTG peuvent être interprétées selon 3 différents cadres de lecture aboutissant soit à des répétitions poly-alanine, soit des répétitions poly-leucine, ou encore être dans la région 3'UTR du gène (voir (Margolis *et al.* 2019) pour revue). Les répétitions CTG ont été associées à la maladie, cependant, leur rôle exact dans la pathogénèse demeure inconnu. Différentes hypothèses sont proposées : (1) perte de fonction de la junctophiline 3 par séquestration dans le noyau des ARN sous forme de *foci* avec (2) gain de fonction toxique des ARNm mutants à l'instar de ce qui est décrit dans la DM1 (voir chapitre suivant), (3) gain de fonction toxique des protéines allongées poly-glutamine ou poly-leucines.

II.3.4.2 LA DYSTROPHIE MYOTONIQUE DE TYPE 1

DESCRIPTION CLINIQUE

La Dystrophie myotonique de type 1, aussi appelée DM1 ou encore maladie de Steinert, est une pathologie neuromusculaire autosomique dominante, décrite au début du XXème siècle par le docteur Hans Steinert (Steinert 1909). Avec une prévalence estimée de 1/20 000 dans le monde, c'est la maladie neuromusculaire la plus fréquente chez l'adulte (Theadom *et al.* 2014). La DM1 présente une extrême variabilité phénotypique ainsi qu'un phénomène d'anticipation très marqué. Le phénomène d'anticipation se défini par une aggravation et une apparition de plus en plus précoce des symptômes au fil des générations. Le principal

symptôme de la DM1 est la myotonie (trouble du tonus musculaire correspondant à des difficultés de relaxation après une contraction volontaire). Cependant, la DM1 est multisystémique, et provoque aussi par exemple une dégénérescence neuromusculaire progressive, des troubles neurologiques, digestifs, des anomalies ophtalmique et cardiaque, ou encore une insulino-résistance. La DM1 se présente selon quatre formes différentes suivant l'âge d'apparition des symptômes (cf. Figure 4 et (Bird 2018) pour revue) :

-<u>La forme tardive</u> passe souvent inaperçue. En effet, les patients développent une faiblesse musculaire et une cataracte bilatérale en écusson assez tardivement. L'espérance de vie de ces patients n'est pas (ou peu) réduite

-La forme adulte (aussi dite « classique ») est multi-systémique avec une myotonie, une dégénérescence des fonctions neuromusculaires, liées à des troubles de la conduction cardiaque. L'on observe également des troubles du comportement, un déficit intellectuel mineur s'aggravant parfois avec l'âge, des défauts endocriniens et respiratoires (pouvant nécessiter une ventilation artificielle). Il a été observé que ces patients ont plus de risque de développer certaines formes de cancers de la thyroïde, du colon, ou encore de la peau par exemple (Emparanza *et al.* 2018, Wang *et al.* 2018)

-<u>La forme juvénile</u> se manifeste avant l'âge de 10 ans. Elle est caractérisée par des troubles cardiaques et neurologiques (troubles du sommeil, de l'attention) ainsi que par une faiblesse musculaire. La forme « juvénile » est létale dans 20% des cas, les sujets survivants ont généralement une espérance de vie réduite (30-40 ans).

-<u>La forme congénitale</u>, est la forme la plus sévère de la maladie. Elle se manifeste dès le stade prénatal, avec des mouvements fœtaux réduits, un poly-hydramnios et souvent avec la formation de pied bot. À la naissance, les enfants présentent une hypotonie sévère, des troubles respiratoires et une diplégie faciale. Le taux de mortalité de ces enfants est très élevé durant les 3 premières années de vie. Les enfants survivants auront une légère amélioration du tonus musculaire durant la prime-enfance, puis leur tableau clinique s'aggrave : ils développent une DM1 classique avec des symptômes sévères accompagné d'un retard mental.

Cependant, la classification des différentes formes de la DM1 est régulièrement sujet à débat. Selon les propositions de classification, on peut distinguer de 4 jusqu'à une quinzaine de formes différentes (Bassez 2019).



Figure 4: Exemple de famille atteinte de la Dystrophie Myotonique de Type 1

DEFAUT MOLECULAIRE ASSOCIE A LA DM1

La DM1 résulte d'une expansion anormale de triplets répétés CTG dans la région 3'UTR du gène *DMPK* (Dystrophy Myotonic Protein Kinase) localisé en q13.3 du chromosome 19. Cette maladie se transmet selon le mode autosomique dominant. Dans la population générale, le nombre de triplets CTG est polymorphe, compris entre 5 et 34 répétitions et reste stable. En revanche, chez les patients DM1, la taille des répétitions est supérieure à 50 CTG et augmente dans 90% des cas au cours des générations, jusqu'à atteindre 4000 répétitions dans la forme la plus grave de la maladie (**instabilité intergénérationnelle**). Les allèles ayant une répétition comprise entre 35 et 49 répétitions sont dits « prémutés » : ces allèles ne provoquent pas la maladie, mais ont de plus grands risques de transmettre une expansion pathologique de la répétition CTG à la génération suivante. Dans les familles DM1, il a été observé une bonne corrélation entre la taille des répétitions et la sévérité des symptômes. Le phénomène d'anticipation trouve ainsi une explication par l'instabilité intergénérationnelle biaisée vers les expansions et la corrélation directe entre la sévérité des symptômes et la taille de l'expansion de triplets CTG (voir (Yum *et al.* 2017) pour revue).

Le nombre de répétitions CTG hérité varie selon la taille de la répétition et le sexe du parent transmetteur. Lorsque le nombre de triplets est <100 CTG chez le parent transmetteur, les plus grandes expansions à la génération suivante sont observées dans le cas de transmissions paternelles. En revanche, au-delà de 500 CTG, on observe une inversion de l'influence du sexe du parent transmetteur. La taille des expansions est plus importante lors des transmissions maternelles et on note de plus fréquentes contractions lors des transmissions

paternelles. Les expansions supérieures à 1000 répétitions (observées le plus souvent chez les patients DM1 atteints de la forme congénitale) sont généralement transmises par la mère. Il a aussi été remarqué, au cours d'études familiales, des phénomènes plus rares, sporadiques ou répétés, de contractions intergénérationnelles dont la fréquence est estimée à 10% dans le cas de transmissions paternelles et seulement à 3% lors de transmissions maternelles (Brunner *et al.* 1993, Martorell *et al.* 2007, Pratte *et al.* 2015).

Chez les patients DM1, la taille des répétitions CTG varie également dans les tissus au cours de la vie du patient (Ashizawa *et al.* 1993). Cette instabilité somatique se caractérise par:

-<u>Une mosaïque inter-tissulaire</u> : c'est-à-dire que tous les tissus n'auront pas le même nombre de répétitions CTG.

-<u>Une mosaïque intra-tissulaire</u> : au sein d'un même tissu, toutes les cellules ne présenteront pas le même nombre de répétitions CTG.

Cette mosaïque est de plus en plus prononcée avec l'âge du patient. L'instabilité somatique est biaisée vers les expansions et est corrélée avec la progression des symptômes DM1.

PHYSIOPATHOLOGIE

Les mécanismes par lesquels l'expansion des répétitions CTG entraîne des troubles multi-systémiques chez les patients sont nombreux et résumés dans la figure 5. *DMPK* code pour une sérine-thréonine kinase, détectée notamment dans le cœur, les muscles squelettiques et les jonctions neuro-musculaires. À l'échelle cellulaire, DMPK a été localisé dans le réticulum sarcoplasmique et les gap-junctions des cellules cardiaques. DMPK est impliquée dans l'intégrité de l'enveloppe nucléaire des myoblastes (Kaliman *et al.* 2008) et interagit avec d'autres kinase influençant le métabolisme de protéines GTPase de la famille Rho et la mort cellulaire médiée par le complexe hexokinase II-Src (Schiavon *et al.* 2002, Kaliman *et al.* 2008, Pantic *et al.* 2013). De plus, il a été démontré chez la souris et le poulet que DMPK participe, au cours de l'embryogénèse, à la régulation de l'apoptose, et serait impliqué dans l'expression de certains marqueurs des cellules musculaires comme la myogénine (Harmon *et al.* 2008). Un modèle murin knock-out pour *DMPK*, a permis de démontrer que DMPK est aussi impliqué dans l'homéostasie insulinique (Llagostera 2007) et calcique (Pall *et al.* 2003).

Chez les patients, l'expression de *DMPK* et *SIX5* (gène dans le voisinage de *DMPK*, en 3' des expansions CTG) est réduite (Groenen *et al.* 1998). Ainsi, les premières hypothèses pour tenter d'expliquer le rôle des répétitions CTG mutantes dans la pathogénèse de la DM1 se sont orientées vers un effet en *cis* de la mutation. Dans ce modèle, les répétitions CTG sont responsables de l'altération de l'expression du gène *DMPK* et *SIX5*, et cette diminution serait

la cause de la maladie. Cependant, les différents modèles murins invalidés pour le gène *Dmpk* n'ont développé qu'une légère myopathie et peu anomalies cardiaques, voire aucun symptôme (Jansen *et al.* 1996, Berul *et al.* 1999, Carrell *et al.* 2016); tandis que les modèles murins invalidés pour *SIX5* (gène impliqué dans l'organogénèse puis exprimé dans la sclère et la rétine adulte) ne présentent que de faibles atteintes cardiaques et une cataracte aux caractéristiques différentes de celles observées chez les patients DM1 (Klesert *et al.* 2000, Wakimoto *et al.* 2002). Donc, l'altération de l'expression des gènes *DMPK* et *SIX5* par l'expansion de la répétition CTG ne semble pas capable à elle seule d'expliquer la totalité des symptômes observés chez les patients. C'est pourquoi la possibilité d'un effet en *trans* a été suggérée.



Figure 5: Représentation schématique de la pathogénèse de la DM1. Les répétitions CTG ont un effet en cis qui diminuent l'expression de DMPK et SIX5. Les transcrits mutants (majoritairement sens, mais aussi antisens) s'accumuleront dans les noyaux en formant des foci dérégulant des protéines essentielles au métabolismes d'autres ARNm (CELF, MBNL, HRNP), ce qui permet de qualifier la DM1 de splicéopathie.

Des études ont montré que les ARNm *DMPK* porteurs d'amplification anormale de rCUG (transcrits sens) et rCAG (transcrits antisens) s'accumulent dans le noyau des cellules de patients DM1 sous forme de foyers d'accumulation nommés « *foci* » et deviennent toxiques pour la cellule (Taneja *et al.* 1995, Mankodi *et al.* 2000, Seznec *et al.* 2001, Michel *et al.* 2015). Les *foci* nucléaires ont un effet en *trans*, avec un gain de fonction des ARNm *DMPK* mutants. Des études ont ainsi démontré qu' il y avait une augmentation de l'activité et de la quantité des

protéines CUG-Binding Protein (CUG-BP, protéines impliquées dans l'épissage alternatif des ARNs) dans les lymphoblastes, les muscles squelettiques, le coeur et le cerveau des patients DM1 (Timchenko et al. 1996, Savkur et al. 2001, Timchenko et al. 2001, Kuyumcu-Martinez et al. 2007). Par la suite, il a été démontré que les protéines de la famille muscleblind (MBNL1, MBNL2 et MBNL3), elles aussi impliquées dans l'épissage alternatif, sont séquestrées au niveau des foci d'ARNm mutants dans des fibroblastes de patients DM1 (Fardaei et al. 2002). Cette séquestration diminue le pool libre nucléocytoplasmique de ces protéines aboutissant à la diminution de leur activité normale. Chez les patients DM1, des défauts d'épissage d'un certain nombre d'ARN ont été effectivement observés (voir (Nakamori et al. 2013, Chau et al. 2015, Wang et al. 2018) pour revue). C'est le cas, entre autres, des transcrits codant pour CLCN-1 (Chloride chanel-1), IR (Insulin Receptor) ou encore des transcrits codant pour le complexe dystrophine-glycoprotéine (DGC) permettant la cohésion des filaments d'actine dans les muscles (Charlet-B et al. 2002, Wheeler et al. 2007, Nakamori et al. 2008). Les découvertes des anomalies d'épissage des gènes IR et CLCN-1 expliquent respectivement au moins en partie les anomalies du métabolisme du glucose et la myotonie caractéristiques de la DM1. Les protéines MBNL et CUGBP sont aussi impliquées dans d'autres métabolismes des ARNm tels que la stabilité, la polyadénylation régulée au cours du développement (Batra et al. 2014), l'adressage et la traduction, la régulation des miARN, la fonctionnalité de protéines : ces mécanismes sont aussi perturbés dans la DM1 (Sicot et al. 2011). Des exemples d'éléments dérégulés dans la DM1 sont résumés dans le tableau 5.

En 2001, la découverte de l'anomalie génétique à l'origine de la dystrophie myotonique de type 2 (DM2), très semblable cliniquement parlant à la DM1, a conforté l'hypothèse d'un effet *trans*-dominant dans les dystrophies myotoniques (DM1 et DM2). La DM2 est causée par l'expansion d'un quadruplet CCTG dans l'intron 1 du gène *CNBP (ZNF9)* (Day *et al.* 2005). Les ARN mutés *ZNF9* s'accumulent sous forme de *foci* nucléaires (comme pour les ARNm mutés *DMPK* dans la DM1). Comme dans la DM1, ces agrégats séquestrent les protéines de la famille MBNL, et accroissent la quantité de protéines CUGBP, ce qui entrainera des défauts de l'expression et d'épissage alternatif d'autres ARNs. La DM1 et la DM2 sont dues à des mutations sur des gènes différents (*DMPK* et *ZNF9*) à fonctions différentes (kinase et facteur de transcription à doigt de zinc), cependant, les répétitions de trinucléotides dans la DM1et la DM2 produisent des ARN riches en CUG. Ainsi, il a été conforté que les similitudes entre ces 2 pathologies étaient liées à l'effet en *trans* des ARN mutants dans la DM1 et la DM2. Toutefois, l'effet *cis* n'est pas exclu et pourrait expliquer les différences entre les symptômes de la DM1 et la DM2 comme, par exemple, la forme congénitale, qui est absente dans la DM2.

Protéine	Rôle des protéines	Type de dérégulation	Conséquences/ Symptômes (démontrés ou supposés)	Références
APP	Récepteur transmembranaire	Expression isoforme aberrant	ND	(Jiang <i>et al.</i> 2004)
BIN1	Invaginations des membranes dans la biogénèse des muscles	Expression isoforme aberrant	Faiblesse musculaire, diminution du nombre de tubules T	(Fugier <i>et al.</i> 2011)
CACNA1	Canal calcique impliqué dans la polarisation membranaire des cellules cardiaques	Surexpression	Problème de conduction de l'influx cardiaque	(Rau <i>et al.</i> 2011)
CACNA1S	Canal calcique impliqué dans le couplage excitation/contraction des muscles squelettiques	Expression isoforme aberrant	Faiblesse musculaire, troubles d'excitation/contraction musculaires	(Tang <i>et al.</i> 2012)
CELF	Régulation de l'épissage alternatif	Hyper-phosphorylation / stabilisation	Expression aberrantes d'isoformes fœtales des ARNm CELF- dépendants	(Wang <i>et al.</i> 2007)
CLCN1	Canal chlore membranaire	Expression isoforme aberrant	Hyperexcitabilité membranaire des muscles, Myotonie	(Charlet-B <i>et al.</i> 2002, Mankodi <i>et al.</i> 2002)
DDX5	Hélicase ARN	Séquestration dans les <i>foci</i> nucléaires de de <i>DMPK</i> mutants, perte de fonction	Favorise la séquestration de MBNL1 dans les <i>foci</i>	(Laurent <i>et al.</i> 2011)
DMD	Protéine du cytosquelette, maintien de la structure et de la signalisation du sarcolème	Expression isoforme aberrant	Dystrophie musculaire, fragilité des membranes des fibres musculaires	(Nakamori <i>et al.</i> 2007)
GJA1	Canal impliqué dans la conductance intra- cardiomyocitaire	Surexpression	Arrythmie	(Groh <i>et al.</i> 2008, Souidi <i>et al.</i> 2018)
GLT1	Transporteur du glutamate	Sous-expression	Perte de coordination des mouvements	(Sicot <i>et al.</i> 2017)
GRIN1	Récepteur au glutamate, impliqué dans la plasticité synaptique	Sur-expression d'un isoforme (x3)	Trouble de la mémoire et de l'apprentissage	(Jiang <i>et al.</i> 2004)
GSK3β	Kinase du métabolisme du glycogène	Suractivé	Faiblesse musculaire et myotonie	(Jones <i>et al.</i> 2012)
hnRNP H	Régulateur d'épissage alternatif	Surproduction	Dérégulation de l'épissage alternatif d'autres gènes	(Paul <i>et al</i> . 2006)
IR	Récepteur à l'Insuline	Prédominance isoforme IR-A à réponse insulinique faible	Insulino-résistance	(Savkur <i>et al.</i> 2001)
LDB3	Composant des bandes Z du muscle cardiaque	Expression isoforme aberrant	Désorganisation des sarcomères, interruption des bandes Z	(Machuca-Tzili <i>et al.</i> 2006)

 Tableau 5: Exemples de protéines dérégulées dans la DM1, conséquences attribuées ou supposées dans la physiopathologie de la DM1

Protéine	Rôle des protéines	Type de dérégulation	Conséquences/ Symptômes (démontrés ou supposés)	Références
MBNL	Protéine en doigt de zinc modulant l'épissage alternatif	Séquestration dans les <i>foci</i> nucléaires de de <i>DMPK</i> mutants, perte de fonction	Dérégulation de l'épissage d'autres ARNm	(Miller <i>et al.</i> 2000)
NKX2.5	Facteur de transcription impliqué dans la conduction cardiaque	Expression aberrante hors du tissus cardiaque	Dérégulation de l'expression des cibles de NKX2.5	(Yadava <i>et al.</i> 2008)
РКМ	Enzyme produisant du pyruvate lors de la glycolyse	Expression isoforme aberrant à activité constitutive	Favorisation de la production de lactate	(Christofk <i>et al.</i> 2008)
RAB3A	Protéine de vésicules synaptiques permettant la fusion des membranes vésiculaires et cellulaire	Surexpression	Trouble comportemental, cognitif et neuropshychologiques	(Hernández-Hernández <i>et al.</i> 2013)
RYR1	Récepteur à la Ryanodine permettant le relargage de calcium des réticulums endo/sarco plasmiques	Expression d'isoforme aberrant	Dégénération musculaire	(Kimura <i>et al.</i> 2005)
SCN5A	Sous unité du canal sodium dépendant impliqué dans l'excitabilité des cardiomyocytes et la conduction de l'influx cardiaque	Expression d'isoforme aberrant	Arythmie, trouble de la conduction cardiaque	(Freyermuth <i>et al.</i> 2016)
SERCA1	ATPase des réticulums endo/sarcoplasmiques	Expression d'isoforme aberrant	Dégénération musculaire	(Kimura <i>et al.</i> 2005)
SHARP	Facteur de transcription	Délocalisation dans le cytoplasme	Sous expression des cibles de SHARP	(Dansithong <i>et al.</i> 2011)
SLITRK 4	Protéines impliquées dans la synaptogénèse et la neuritogénèse	Sous expression	Troubles du développement des connexions neuromusculaire et la formation des neurites	(Marteyn <i>et al.</i> 2011)
STAU1	ARN double brin impliqué dans l'adressage des ARNm dans les différents compartiments subcellulaires	Surexpression	Favorise la localisation nucléaire des ARNm <i>DMPK</i> mutant et dérégule le splicing d'autres ARNm	(Ravel-Chapuis <i>et al.</i> 2012)
SYN1	Régulation du relargage de neurotransmetteurs synaptique via des signaux de phosphorylation	Hyperphosphorylation	Trouble comportemental, cognitif et neuropsychologiques	(Hernández-Hernández <i>et al.</i> 2013)
TAU (MAPT)	Protéine associée aux microtubules, stabilisation du cytosquelette des axones	Expression d'un isoforme aberrant	Enchevêtrement neurofibrillaire, axonopathie	(Sergeant <i>et al.</i> 2001, Jiang <i>et al.</i> 2004, Caillet- Boudin <i>et al.</i> 2014)
TNNT2	Protéine du complexe troponine cardiaque	Expression d'isoforme aberrant	Hyper sensibilité des myofibrilles au calcium	(Philips <i>et al.</i> 1998)

Tableau 5: Exemples de protéines dérégulées dans la DM1, conséquences attribuées ou supposées dans la physiopathologie de la DM1

II.3. CONCLUSION DE CHAPITRE

Les différentes maladies présentées dans ce chapitre découlent de répétitions de différentes natures, sur différentes régions géniques. Ces répétitions ont des conséquences variées au niveau des ARN et/ou des protéines. Pourtant, la plupart de ces maladies partagent des mécanismes communs d'instabilité, véritables piliers de ces pathologies. La compréhension des mécanismes d'instabilité des TNRs est primordiale pour comprendre l'origine et la mécanistique de la DM1 ainsi que des autres maladies à expansion de triplets instables. De nombreuses études dans différents modèles *in vitro*, *in vivo* et chez les patients ont permis de mieux comprendre ces mécanismes. La suite de mon introduction sera consacrée aux connaissances accumulées sur les différents mécanismes et facteurs.

III- INSTABILITE DES TRIPLETS REPETES

Il existe de nombreux facteurs influençant l'instabilité des triplets répétés, à différents stades de la vie des patients. Nous décrirons ici dans un premier temps les deux grands types d'instabilités (intergénérationnelle et somatique), puis l'influence de la nature des répétitions sur l'instabilité. Enfin, nous présenterons les différentes voies du métabolisme de l'ADN qui viennent moduler l'instabilité.

III-1. INSTABILITE INTERGENERATIONNELLE

III.1.1. INFLUENCE DU PARENT TRANSMETTEUR : EVENEMENTS GERMINAUX

L'influence du sexe du parent transmetteur varie en fonction des pathologies. À titre d'exemple, dans la HD, les transmissions paternelles transmettent les plus longues expansions tandis les transmissions maternelles témoignent plus rarement d'une expansion (Kremer *et al.* 1995). A l'inverse, les plus longues expansions de triplets CTG dans la DM1 sont observées dans le cas de transmission maternelles (Brunner *et al.* 1993). On peut aussi citer l'exemple de la FRDA, où les expansions de triplets GAA ne sont observées que dans les transmissions maternelles, alors que les transmissions paternelles résultent en des contractions (De Michele *et al.* 1998). Ces observations laissent supposer que l'instabilité peut intervenir dès la gamétogenèse. La figure 6 ci-dessous apporte un bref rappel sur les spécificités des gamétogénèse mâle et femelle, afin de mieux appréhender en quoi les différences entre ces 2 évolutions cellulaires sont importantes pour expliquer l'influence du sexe du parent transmetteur dans l'instabilité intergénérationnelle.



Figure 6: Rappel sur les processus d'ovogénèse et spermatogénèse chez l'Homme.

III-1.1.1. L'INSTABILITE DE LA GAMETOGENESE MALE

Les premières informations sur les mécanismes d'influence de l'instabilité intergénérationnelle chez les mammifères ont été rendue accessibles grâce au développement de modèles murins. En effet, il a été montré dans différents modèles murins que l'instabilité de triplet CAG (modèle de maladie de Huntington) et CTG (modèle de la DM1) dans les lignées germinales mâles et femelle était hautement dépendante du MMR, notamment de *Msh2* (facteur primordial du MMR, voir chapitre III.4.2.3. pour plus de détails). Par exemple, il a été démontré que *Msh2* était nécessaire pour observer l'instabilité des triplets CAG dans un modèle murin invalidé pour *Msh2* (*Manley et al. 1999*). Dans un modèle murin de la DM1, en cas d'invalidation de *Msh2*, il y a une disparition des expansions au profit de la formation de contractions. Dans ce même modèle murin développé par le laboratoire, il a été montré que l'instabilité des triplets CTG intervenait dans les gamètes (spermatogonies) mais également aux premiers stades du développement de l'embryon (Savouret *et al.* 2003, Savouret *et al.* 2004). L'étude des modèles murins a aussi démontré que *Msh3* était nécessaire et limitant pour les expansions intergénérationnelles, (Wheeler *et al.* 2003, Foiry *et al.* 2006, Dragileva *et al.* 2009).

Chez les patients atteints de la chorée de Huntington, les évènements d'expansion de triplets CAG sont majoritairement observables dans les stades pré-méiotiques, bien qu'ils puissent aussi plus rarement se manifester après la méiose dans les spermatozoïdes (Jansen *et al.* 1994, Yoon *et al.* 2003). Pour les patients HD, l'âge du père transmetteur ne semble pas influencer la taille de la répétition héritée par le fœtus (Martorell *et al.* 2004, Wheeler *et al.* 2007). En revanche, dans le cas de la DM1, il existe une forte corrélation entre l'âge du père transmetteur, la taille des répétitions présentes dans les spermatozoïdes et la taille de répétition héritée (Morales *et al.* 2015). L'ensemble de ces études laisse supposer que la majorité des évènements d'instabilité de TNRs dans la lignée germinale mâle se produit dans les stades précoces de la gamétogénèse, qui sont des phases de mitoses et réplication intensives.

Dans le cas d'une transmission paternelle de la DM1, les plus longues expansions dans les gamètes sont observées chez les patients porteurs de pré ou proto-mutation (Martorell *et al.* 2000). Il a aussi été prouvé que lorsqu'un homme hérite d'une grande taille de répétition CTG, la mosaïque somatique dans sa lignée germinale est biaisée vers les contractions (Jansen *et al.* 1994). Ces données suggèrent l'existence d'un mécanisme de protection contre les larges expansions au cours de la spermatogénèse. Ceci expliquerait pourquoi les plus grandes tailles de répétitions, menant notamment à la forme congénitale de la DM1 ne sont quasiment qu'exclusivement héritées *via* les transmissions maternelles, au cours desquelles

l'instabilité des triplets dépendrait de règles différentes (Harley *et al.* 1993, Lavedan *et al.* 1993, Ashizawa *et al.* 1994, Barcelo *et al.* 1994, Barbé *et al.* 2017)

III-1.1.2. INSTABILITE DE LA GAMETOGENESE FEMELLE

Au cours de l'ovogénèse humaine, l'instabilité des TNRs est active dans les stades pré-méiotiques dans les cas de syndrome de l'X fragile et de la DM1 (Oberle *et al.* 1991, Yu *et al.* 1991, De Temmerman *et al.* 2004, Dean *et al.* 2006). Contrairement à la ligné germinale mâle, il a été observé, une instabilité biaisée vers les expansions dans les ovocytes au cours de la vie des patientes atteintes de la DM1 et du FXS (Yrigollen *et al.* 2014, Morales *et al.* 2015). Sachant que les ovocytes sont bloqués à la Prophase I à la naissance, ceci signifierait que l'instabilité intervient dans des cellules non mitotiques. Donc, contrairement aux lignées germinales paternelles, l'instabilité dans les ovocytes ne dépendrait pas majoritairement de la réplication, mais de d'autres mécanismes modulateurs, comme par exemple la réparation de l'ADN dans les ovocytes à l'arrêt en méiose I.

L'instabilité dans les cellules germinales représente la première vague d'instabilité intergénérationnelle qui sera suivie par d'autres au cours du développement du fœtus.

III-1.2. INSTABILITE AU COURS DU DEVELOPPEMENT : ÉVENEMENTS ZYGOTIQUES

Il a été observé des cas de jumeaux monozygotiques atteints du syndrome de l'X fragile ayant hérité de tailles de répétitions identiques, mais ayant à la naissance des tailles de répétitions différentes (Devys *et al.* 1992, Helderman-van den Enden *et al.* 1999, Reyniers *et al.* 1999). Toujours dans les cas d'X fragiles, les hommes qui présentent une mutation complète n'ont dans leurs gamètes que des tailles de répétitions inférieures à celle héritée (Malter *et al.* 1997). Ces éléments suggèrent que l'instabilité des répétitions CGG dans le cas du FXS démarre très tôt au cours de l'embryogénèse. Au contraire, dans le cas de la DM1, les répétitions CTG restent stables durant le premier trimestre de gestation, et l'instabilité n'est détectée qu'entre la 13e et 16e semaine de la gestation (Martorell *et al.* 1997). Par ailleurs, il a également été démontré chez des fœtus humains DM1, que lorsque le père transmetteur est porteur d'une pré-mutation, le nombre de répétitions CTG peut s'accroître au cours de l'embryogénèse (Dean *et al.* 2006). Par ailleurs, dans le cas de patients HD, l'instabilité des répétitions CAG au cours du développement varie en fonction du sexe du foetus lors de transmissions maternelles : les enfants de sexe féminin naissent avec une taille de répétitions plus élevée que les garçons (Wheeler *et al.* 2007).

Ces différents éléments mettent en évidence que selon les pathologies à TNRs, l'instabilité zygotique est différente, ce qui laisse suggérer que la régulation et/ou la nature des mécanismes impliqués dans l'instabilité des TNRs sont différentes.

L'instabilité intergénérationnelle est la résultante de 2 vagues d'instabilités : (1) au cours de la gamétogénèse et (2) au cours du développement du fœtus. Ces évènements conditionneront la taille de la répétition de triplets héritée. Après la naissance, l'instabilité des triplets se poursuit au cours de la vie des patients, on parle alors d'instabilité somatique.

III.2. INSTABILITE SOMATIQUE

Généralement, l'instabilité somatique est biaisée vers les expansions et est tissu spécifique. C'est-à-dire qu'au sein d'un même individu, les différents tissus ne vont pas présenter la même mosaïque de TNRs. De manière générale, l'amplitude de l'instabilité somatique augmente avec l'âge du patient. Ceci en est particulièrement le cas dans la DM1 ou il existe une bonne corrélation entre l'avancée des symptômes et la progression de l'instabilité au cours de la vie des patients (Overend *et al.* 2019).

Comme pour l'instabilité intergénérationnelle, l'instabilité somatique va présenter des particularités différentes en fonction des pathologies. Par exemple, l'instabilité est continue au cours de la vie des patients DM1 dans différents tissus (cerveau, muscle, sang, coeur...), alors que dans la HD, SCA1, et DRPLA, l'instabilité somatique n'est présente que dans le cerveau (voir (Pearson *et al.* 2005, López Castel *et al.* 2010, Dion 2014) pour revue). L'instabilité somatique se manifeste donc dans des cellules quiescentes et prolifératives, et indépendamment de leur degré de différenciation (Ku *et al.* 2010, Du *et al.* 2013). Ces observations laissent supposer que l'instabilité des triplets dépend de différents mécanismes, qu'ils soient propres aux répétitions en elles-mêmes ou bien acteurs du métabolisme de l'ADN (Réplication, Réparation et Transcription).

III.3. FACTEURS D'INFLUENCE DE L'INSTABILITE

III.3.1. LA PURETE DES REPETITIONS

La première corrélation entre une interruption et la stabilisation des répétitions a été faite dans le cadre de la SCA1, au cours d'une étude sur une centaine de patients qui a révélé que la majorité des répétitions CAG des allèles sains sont interrompus par des triplets CAT. Selon les auteurs, la perte de ces interruptions serait responsable de l'apparition des tailles de répétitions CAG qui franchissent alors un seuil pathologique (Chung *et al.* 1993). Cette hypothèse est soutenue par les auteurs qui ont également mis en évidence que lorsque des

patients porteurs de répétitions pathologiques présentent aussi des interruptions CAT, l'instabilité somatique était réduite chez ces patients. De plus, une autre étude a démontré que les interruptions CAT dans la SCA1 étaient associées à des contractions intergénérationnelles chez les patients (Menon *et al.* 2013). La même observation a été faite dans des familles atteintes par d'autres pathologies, comme dans la FRDA avec des interruptions GAAAGAA au sein des répétitions GAA, associées à une stabilisation intergénérationnelle des triplets (Cossée *et al.* 1997). Dans le syndrome de l'X fragile, la présence d'interruptions AGG a été associée à une stabilisation des répétitions CGG (Latham *et al.* 2014); il en va de même pour les interruptions CAA dans les répétitions CGG dans le cadre de la SCA2 (Choudhry *et al.* 2001). Cependant, dans le cas de la SCA8, aucune des interruptions CCA, CTA, CTT, CTC et CCG n'a été associée à une stabilisation des répétitions chez les patients (Moseley *et al.* 2000).

Dans le cadre de la DM1, plusieurs types d'interruptions ont été identifiées chez les patients, le tableau 1 en annexe résume les interruptions trouvées dans différentes familles DM1, dont celles étudiées au laboratoire et discutées dans ce manuscrit. Brièvement, il a été identifié dans différentes cohortes de patients DM1, des interruptions CCG, CTC et GGC avec une prévalence comprise entre 3-7% (Musova et al. 2009, Braida et al. 2010, Santoro et al. 2013, Cumming et al. 2018). Les interruptions CCG et GGC sont transmises indépendamment du sexe du parent transmetteur, mais, il est intéressant de noter que les interruptions CTC sont majoritairement relevées de novo, suite à une transmission paternelle de la DM1. À ce jour, il est théorisé que les interruptions CTC se produisent suite à l'instabilité post-natale dans la lignée germinale mâle où la réplication reste intensive (Cumming et al. 2018). Les interruptions ont été identifiées aussi bien en en 5' qu'en 3' des répétitions, au sein d'une large amplitude de taille de répétions saines et pathologiques : entre 37 et plus de 1000 CTG (P.Leeflang et al. 1995, Musova et al. 2009, Botta et al. 2017). Dans toutes ces études dans la DM1, il a été observé que la mosaïque somatique était de moindre amplitude en cas d'interruptions, et que l'instabilité intergénérationnelle était généralement stabilisée voire parfois biaisée vers les contractions (Botta et al. 2017, Pešović et al. 2017).

De manière intéressante, les interruptions et la stabilisation des répétitions liées aux interruptions ont été associées à une pénétrance réduite des pathologies : les interruptions CAT dans la SCA1 sont associées à des symptômes moins sévères (Menon *et al.* 2013) ainsi que dans le Syndrome de l'X fragile (Eichler *et al.* 1994, Kunst *et al.* 1994). Dans la DM1 le phénomène d'anticipation est parfois absent dans le cas de transmission des interruptions (Musova *et al.* 2009, Cumming *et al.* 2018, Pešović *et al.* 2018). En revanche, dans la SCA8 tout comme aucune stabilisation n'est observée en présence d'interruptions, aucune réduction de la pénétrance de la pathologie n'est observée (Moseley *et al.* 2000).

Les mécanismes exacts par lesquels les interruptions stabiliseraient les répétitions et réduiraient l'avancée des pathologies à TNRs demeurent peu connus. Les interruptions pourraient stabiliser les répétitions en changeant la structure secondaire imposée par les répétitions, et pourraient servir de point d'ancrage/repérage évitant les glissements et mésappariements des brins d'ADN lors de la réplication, transcription et réparation, qui sont des processus sources d'instabilité comme nous le verrons dans la suite de ce manuscrit (Pearson *et al.* 1998, Choudhry *et al.* 2001). Ou encore, la survenue d'interruption(s) diminuerait le nombre de répétitions continues, parfois jusqu'en dessous du seuil pathologique, expliquant ainsi la stabilisation des TNRs en cas d'interruption(s).

III.3.2. LA LONGUEUR DES REPETITIONS

La longueur des répétitions est le premier facteur ayant été corrélé à la sévérité et la progression des symptômes des maladies à triplets. C'est également un facteur primordial de la dynamique de l'instabilité des triplets répétés. Pour la plupart des maladies à TNRs, en dessous d'un seuil d'environ 30-40 répétitions, les triplets sont stables. Ce n'est qu'une fois ce seuil franchi, que les répétitions deviennent instables au cours des générations et dans les tissus.

Le lien entre la taille des répétitions et l'instabilité des triplets a été observé chez l'Homme et dans différents modèles (voir (Pearson *et al.* 2005, López Castel *et al.* 2010) pour revues). Par exemple, dans le modèle murin de la DM1 du laboratoire, aucune instabilité n'est détectée lorsque le nombre de répétitions est égal à 20 CTG, une instabilité modérée est détectable chez les souris avec 55 CTG répétés, tandis que, chez les souris avec 300 répétitions, l'instabilité des triplets est très marquée (Gourdon *et al.* 1997, Seznec *et al.* 2000), le même constat a été fait dans différents modèles murins du FXS, FRDA et de la HD (Peier *et al.* 2002, Al-Mahdawi *et al.* 2004, Dandelot *et al.* 2017) pour revue), dans différents modèles de levure (Freudenreich *et al.* 1998, Rolfsmeier *et al.* 2001, Dixon *et al.* 2004, Shishkin *et al.* 2009, Cherng *et al.* 2011) et de cellules humaines DM1 et FXS (Burman *et al.* 1999, Farrell *et al.* 2006, Ditch *et al.* 2009, Du *et al.* 2013).

La raison pour laquelle les répétitions ne deviennent instables qu'une fois un seuil franchi n'est pas encore complètement comprise. Parmi les explications possibles, il a été étudié la formation de structures secondaires imposées par les répétitions anormales.

III.3.3. LES STRUCTURES SECONDAIRES

Théoriquement, toutes les répétitions microsatellites peuvent causer des décalages entre les 2 brins de la double hélice d'ADN (slipped strand). Ces décalages mènent à la formation de structures secondaires ne correspondant pas au repliement classique en double hélice ß. Les longues séquences répétées favorisent le super-enroulement de l'ADN, ce qui mène à des structures secondaires non canoniques thermodynamiquement plus stables que le repliement classique en double hélice ß. La plupart de ces structures secondaires présentent des zones simple brin, des mésappariements et sont maintenues par des liaisons hydrogène. La nature, la pureté et la longueur des répétitions influencent la formation de ces structures secondaires.

Il a très tôt été démontré que les répétitions CTG et CGG modifiaient la structure secondaire de l'ADN (Chastain *et al.* 1995). Depuis, différents types de structures secondaires ont été identifiées selon la nature des répétitions. La suite de ce chapitre propose un aperçu des principales structures identifiées dans les TNRs impliquées dans des pathologies à mutation dynamique.

III.3.3.1. HAIRPINS

Les hairpins (ou « structures en épingle à cheveux ») se forment dans des séquences répétées inversées. Les hairpins sont constituées d'une « tige » avec des bases appariées et d'une boucle simple brin à son sommet. Ces structures peuvent se former via 2 mécanismes. Lorsque l'ADN est double brin, les appariements inter-brins sont « convertis » en appariement intra-brin. En effet, cette structure est thermodynamiquement plus favorable dans les zones à répétitions, elles sont donc plus stables que la structure classique de l'ADN (Gacy *et al.* 1995, Gacy *et al.* 1998). Une autre possibilité est que lorsque l'ADN se retrouve à l'état simple brin (au cours de la réplication, réparation, voire dans le cas de certaines infections virales), celuici se replie sur lui-même, formant des hairpins. Il est aussi possible qu'au moment du réappariement, il y ait un glissement des brins (à cause de la monotonie des répétitions), aboutissant au repliement aberrant des triplets répétés pour forcer le ré-appariement des séquences flanquantes. S'il y a formation d'hairpin sur les 2 brins complémentaires, l'une en face de l'autre, on parlera alors de structures cruciformes (voir figure 7).

La plupart du temps, les hairpins ont été observées *in vitro* avec des oligonucléotides ou des plasmides grâce à une analyse par migration sur gel ou par techniques de type analyse de point de fusion ou dichroïsme circulaire. Cependant une expérience en culture cellulaire humaine se basant sur un clivage spécifique a confirmé la présence d'hairpin CTG/CAG *in vivo (Liu et al. 2010)*.



Figure 7: Schématisation d'hairpins composées de répétitions CTG.

Les répétitions (CCG)n, (CGG)n, (CTG)n et (CAG)n sont connues pour former des hairpins au-delà de 4 répétitions (Mitas *et al.* 1995, Chi *et al.* 2005, Liu *et al.* 2010). Plus la répétition est longue, plus la formation d'hairpin stables est probable (Gacy *et al.* 1998). La stabilité des hairpins dépendra aussi des mésappariements dans leur tige. Ainsi les hairpins composées de répétitions CGG sont plus stables que celles composées de répétitions CCG et CTG, qui sont toutes deux équivalentes et plus stables que les hairpins composées de CAG (Gacy *et al.* 1995). Cependant, cette stabilité est hautement dépendante des conditions environnementales où se trouve l'ADN. En effet, en cas de fort encombrement moléculaire (comme lors de la transcription par exemple), cet ordre est modifié pour donner le classement suivant par ordre décroissant de stabilité : hairpin CTG ≥ hairpin CGG > hairpin CAG > hairpin CCG. Ainsi, dans le cas des maladies à triplets, les hairpins formées sur les brins sens et antisens n'auront pas la même stabilité, et cette stabilité variera aussi en fonction de l'activité et cycle cellulaire (Teng *et al.* 2018).

Des études *in vitro* ont permis de mettre en évidence que les hairpins étaient naturellement résolues par des hélicases spécifiques et le complexe RecA/SSB au cours de la réplication chez la bactérie, mais que cette voie se dérégulait avec l'augmentation du nombre d'hairpin à traiter sur une même séquence (Reddy *et al.* 2000). Ainsi, les répétitions anormalement longues, qui forment plus fréquemment des hairpins, déréguleraient ce système de résolution, favorisant ainsi la stabilité des hairpins. Une autre expérience à partir

Les répétitions CTG impliquées dans la DM1 sont connues pour former des hairpins. (a.) Exemple d'hairpin simple, où les bases CG sont appariées dans la tige, tandis que les T sont mésappariés. La boucle au sommet est simple brin. (b.) Exemple de structure cruciforme : 2 hairpins en face l'une de l'autre se forment sur chacun des deux brins d'ADN.

d'oligonucleotides *in vitro* a démontré que des protéines du mécanisme de réparation de mésappariement de base (MMR) était recruté au niveau des hairpins mais que son activité y était inhibée (Owen *et al.* 2005). Les hairpins participeraient donc à une dérégulation du MMR par séquestration et perte d'activité, ce qui favoriserait l'instabilité des répétitions. Ou encore, il a aussi été démontré *in vitro* que les hairpins sont hypersensibles aux mécanismes du BER (Base Excision Repair), avec le recrutement de la protéine OGG1 induisant des cassures au niveau des boucles simples brin du sommet de l'hairpin, notamment lorsqu'une guanine y est présente (Jarem *et al.* 2009).

III.3.3.2. G- QUADRUPLEXES

Les structures G-quadruplexes (GQ ou G4s) reposent sur la formation de tétrades de guanines. Ces structures sont polymorphiques : elles peuvent être intra ou inter caténaires, uni, bi ou tétra-moléculaires, parallèles ou antiparallèles et les boucles générées peuvent être de différentes tailles. La stabilité de ces structures réside en la liaison hydrogène générée à chaque tétrade de guanine qui forment des surfaces planes qui s'auto-empilent les unes sur les autres. Cet empilement est stabilisé par des cations monovalents. La figure 8 ci-après schématise différentes configurations possibles de ces structures.

Il a été démontré tout d'abord *in vitro* et par prédictions bio-informatiques que les G4s se formaient dans les régions simple brin riches en guanine (d'où leur nom), avec une surreprésentation dans les télomères, l'ADN ribosomique, les régions promotrices et 5' UTR. Plus récemment, des études *in vivo* ont confirmé ces prédictions dans des cellules humaines, notamment par immunofluorescence ou cartographie (G4-seq) ou ChIP-seq (voir (Hänsel-Hertsch *et al.* 2017) pour revue et (Sundquist *et al.* 1989, Maizels 2006, Lipps *et al.* 2009)). Il a été démontré *in vitro* ainsi que dans des modèles cellulaires et de *C.elegans* que ces structures résultent notamment de l'action d'hélicases ATP-dépendantes spécifiques au cours de la réplication, ou encore que les G4s feraient suite au recrutement de protéines chaperonnes.



Nature Reviews | Molecular Cell Biology

Figure 8: Schématisation de différentes structure G-quadruplexes.

Les structures G-quadruplex (G4) peuvent être générées à partir d'un seul brin d'ADN (structures unimoléculaires) ou de plusieurs brins d'ADN se rejoignant (par exemple, des structures bi ou tétramoléculaires). Les structures G4 peuvent être classées selon l'orientation des brins : les G4 parallèles ont leurs brins orientés dans le même sens, alors que les G4 antiparallèles ont des orientations de brins en alternance. (a.) Représentations structurelles (à gauche) et schématiques (à droite) d'un G-tétrade qui constitue le noyau des structures G4, qui sont stabilisés par la coordination d'un cation alcalin (orange). (b.) Représentation schématique d'un G4 unimoléculaire parallèle. (c.) Représentation schématique d'un tétra-moléculaire parallèle G4. (d.) Représentation schématique d'une structure G4 antiparallèle unimoléculaire. (e.) Représentation schématique d'une structure G4 unimoléculaire antiparallèle contenant un renflement. (Figure traduite d'après Hänsel-Hertsch, Di Antonio et al. 2017).

Les G4s endossent plusieurs rôles dans les génomes eucaryotes, comme le maintien des télomères ou la régulation de l'expression, et sont déterminants pour l'initiation de la réplication (voir (Rhodes *et al.* 2015, Simone *et al.* 2015) pour revues).

Le cas du Syndrome de l'X Fragile est particulièrement illustratif du rôle des G4s dans l'instabilité des triplets et la pathogénèse. En effet, il été démontré que les répétitions pathologiques de CCG peuvent former des structures G4s en plus des hairpins. La formation de G4s sur ces répétitions pathologiques est d'autant plus stable que la répétition est longue. A l'instar des hairpins, les G4s joueraient un rôle dans l'instabilité des répétitions CGG pathologiques en favorisant le recrutement d'acteurs de la réparation de l'ADN ainsi que la formation d'R-loops, éléments vecteurs de l'instabilité (Zhao *et al.* 2016). De plus, les répétitons CCG pathologiques formant des G4 dans le gène *FMR1* induisent, entre autres, une hyperméthylation invalidant le gène *FMR1*. L'invalidation du gène *FMR1* aboutit à une perte de production de la protéine FMRP (Pieretti *et al.* 1991, Sutcliffe *et al.* 1992). Or, FMRP est une protéine fixant les structures G4s sur les ARNm, empêchant ainsi leur traduction (Feng *et al.* 1997). Une dérégulation de FMRP implique une mauvaise prise en charge de ses ARNm

l'X-Fragile, les G4s sont à la fois (1) une cause d'instabilité des répétitions dont les expansions sont à l'origine du FXS, et (2) sont en partie responsables de la perte de fonction de FMRP, protéine elle-même dévouée au traitement des ARNm porteurs de G4s (Simone *et al.* 2015).

Dans les génomes, il y a des séquences impliquées naturellement (sans être sujette à des mutations) dans la formation de G4s, et ces séquences restent stables, tandis que les répétitions impliquées dans les maladies à triplets et qui forment des G4s sont instables. Ce qui laisse supposer que malgré une configuration *a priori* similaire, les G4s dans les répétitions riches en guanines pathologiques et normales obéissent à des systèmes de régulations différents ou du moins modulés différemment (Weisman-Shomer *et al.* 2000).

III.3.3.3. TRIPLEXES

Les structures triplexes sont composées de 3 brins d'ADN, où le 3^{ème} brin vient s'intercaler dans le grand sillon de la double hélice. Ces structures peuvent être intermoléculaires, dans ce cas le troisième brin provient d'un simple brin d'ADN isolé ou d'une autre molécule d'ADN ; soit intramoléculaire (ou Hinge-DNA, ADN-H), lorsqu'une séquence répétée a une symétrie en miroir, la molécule d'ADN se replie sur elle-même (voir figure 9). Dans le génome humain, les triplexes intra moléculaires sont majoritairement présents dans les régions promotrices et points chauds de la recombinaison. Il a été démontré *in vitro* que ces structures inhibent la transcription et l'élongation (Ohshima *et al.* 1998, Krasilnikova *et al.* 2007), et qu'elles ont un haut potentiel mutagène car ces structures sont reconnues par des systèmes de réparation induisant des cassures (voir (Shah *et al.* 2015) pour revue).



Figure 9: Vue d'ensemble sur la nature des structures triplexes.

(a.) Présentation schématique d'un triplexe à motif de purine intermoléculaire à orientation antiparallèle formée de trois brins d'ADN distincts. (b.) Exemples de structures intramoléculaires ADN-H pouvant se former à partir d'une séquence répétée miroir dans un double brin d'ADN. Les traits noirs représentent des appariements de bases classiques d'ADN double brin en double hélice ß, les points noirs représentent les liaisons hydrogène Hoogsteen inversées par lesquelles la formation triplexe est rendue possible. Les brins bleu et orange représentent les brins de la double hélice classique, le brin vert représente une séquence simple brin extérieure, pouvant provenir d'une autre molécule d'ADN. Figure adaptée et traduite d'après (Holder, Wagner et al.2015)

Les répétitions GAA/TTC (responsables de la FRDA) sont connues pour générer des triplexes en plus d'hairpins (Potaman *et al.* 2004). Il est démontré que la formation de triplexes par les répétitions GAA est en partie responsables de l'inhibition de la transcription de *FXN* (Ohshima *et al.* 1998, Grabczyk *et al.* 2000). Par ailleurs, des études ont démontré *in vitro* que les interruptions GGA/TCC présentées par des patients asymptomatiques, inhibent la formation de triplexes et augmente le niveau de transcription de *FXN* (Sakamoto et al. 2001).

III.3.3.4. DUPLEX UNWINDING ELEMENT (DUE)

Les DUE sont des structures relâchées qui se forment dans la double hélice d'ADN en cas de mésappariement de bases (voir figure 10). Les « renflements » qui se forment peuvent être composés d'un ou plusieurs nucléotides et sont classés en fonction de leur localisation : simple brin, double brin (boucle interne), ou à la jonction entre 2 brins. Ce type de structure se rencontre dans les séquences riches en AT, notamment au niveau des origines de réplication. Ces structures relâchées permettent un accès privilégié à la machinerie de réplication.



Figure 10: Schématisation d'un DUE.

Les répétitions ATTCT/AGAAT riches en AT sont favorables au désenroulement de la double hélice d'ADN, laissant alors se former des régions d'ADN non apparié.

À ce jour, aucune répétition de triplets n'a été associée à ce type de structures. Cela dit, la répétition de ATTCT, responsable de la SCA10 forment ce type de structures dans un modèle *E.Coli (Potaman et al. 2003)*. Dans le cadre de la SCA10, il a été proposé que la répétition ATTCT induise une ré-initiation de la réplication de l'ADN, favorisant l'instabilité des répétitions (Potaman *et al.* 2003, Nedelcheva *et al.* 2005).

III.3.4. LES SEQUENCES ENVIRONNANTES

Plusieurs modèles et études chez les patients ont démontré l'importance des séquences environnantes sur l'instabilité des triplets répétés. Nous décrirons par la suite certains exemples illustrant comment l'environnement génique peut influencer l'instabilité des répétitions impliqués dans les pathologies à mutations dynamiques.

III.3.4.1. L'ENRICHISSEMENT EN GC

Il a été démontré que la **richesse en GC** des séquences environnantes, jusqu'à une distance de 1000pb des répétitions, est positivement corrélée à « l'expandabilité » des répétitions CAG (Brock *et al.* 1999, Nestor *et al.* 2011). Cette observation est cohérente avec la capacité accrue des régions riches en GC à notamment former des structures secondaires non canoniques, à moduler le repliement de la chromatine par la méthylation, ainsi que leur aspect favorable à la formation des R-loops. Or, tous ces éléments sont acteurs de l'instabilité (pour plus de détails, se référer au chapitre III.4.).

III.3.4.2. LES SEQUENCES ENVIRONNANTES LIEES AUX ALLELES MUTANTS

Dans plusieurs maladies à triplets instables certains haplotypes sont en fort déséquilibre de liaison avec les répétitions pathologiques comme dans la maladie de Huntington (Goldberg et al. 1995, Warby et al. 2009, Chao et al. 2017), la SCA2 (Choudhry et al. 2001), la SCA3 (Costa et al. 2019), le Syndrome de l'X Fragile (Zhong et al. 1996, Batra et al. 2014), l'Ataxie de Friedreich (Chauhan et al. 2002) ou encore la DM1 (Imbert et al. 1993, E.Neville et al. 1994, Kumar et al. 2015). Ces observations ont mené à suggérer que (1) les allèles mutants dérivaient d'un ancêtre commun, et (2) qu'il y aurait des allèles prédisposés aux expansions. Par ailleurs, il a été démontré que la nature des séguences environnantes influençait l'instabilité des triplets répétés grâce au développement de différents modèles. Par exemple, le modèle murin DM55 (DM1), généré par micro-injection de répétitions CTG comprises dans 45kpb de séquences environnantes humaines a été le premier modèle DM1 à présenter une instabilité des triplets CTG comparable à celle observée chez les patients (Gourdon et al. 1997). La comparaison de l'instabilité présentée par différents modèles murins de la SCA7, tous porteurs de 92 CAG, mais au sein de différents environnements géniques, a également démontré que l'instabilité des triplets CAG était dépendante des séguences environnantes (Libby et al. 2003). Par ailleurs récemment, une étude chez des familles de patients porteurs du FXS a mis en évidence que certains haplotypes étaient lié à des phénomènes de larges contractions intergénérationnelle (Maia et al. 2017).

III.3.4.3. LOCALISATION DES ORIGINES DE REPLICATION (OR)

Une étude sur le syndrome de l'X Fragile a démontré l'importance de la localisation des origines de réplication dans l'instabilité des triplets CGG. Le gène *FMR1* est normalement répliqué grâce à 2 **origines de réplications** de part et d'autre des répétitions CGG. Il a été démontré qu'un SNP (Polymorphisme d'un Seul Nucléotide) à 50kpb en amont des répétitions CGG, est associé quasiment systématiquement à l'allèle mutant, et qu'il empêche la réplication de démarrer en amont des répétitions (Ennis *et al.* 2007, Gerhardt *et al.* 2014, Gerhardt *et al.* 2014). Dans leur revue, McGinty *et al.* proposent que la réplication ne prend que unidirectionnelle favorise l'instabilité. En effet, dans le cas du FXS, la réplication ne prend que

le brin porteur de CGG comme brin modèle, brin qui forme des hairpins beaucoup plus stables que celles composées de répétitions CCG sur le brin complémentaire : la réplication sera donc potentiellement plus souvent interrompue en présence du SNP lié à l'allèle pathologique, ce qui favoriserait l'instabilité (voir (McGinty *et al.* 2018) pour revue et chapitres III.3.3.1. et III.4.1.). Dans le cas du *locus DMPK*, 2 origines de réplications actives ont été identifiées en cultures cellulaire : une en aval des CTG et une en amont (Cleary *et al.* 2010, Liu *et al.* 2012). Une étude dans un modèle de cellules COS1 transfectées avec des épisomes porteurs des répétitions CTG, a mis en évidence une corrélation entre la distance des origines de réplication et la dynamique de l'instabilité des triplets CTG (Cleary *et al.* 2002). Dans ce modèle, les auteurs avaient la possibilité d'éloigner ou de rapprocher des répétitions CTG une origine de réplication type SV40-ori (voir figure 11 ci-après).



Figure 11: Détermination de la variation de la taille des triplets CTG en fonction du positionnement de l'origine de réplication.

Les schémas en bleu représentent l'orientation et la distance entre l'origine de réplication (SV40) par rapport à la séquence répétée de 79 CTG. Sur les graphiques, les barres blanches représentent la fréquence à laquelle différentes tailles de répétitions étaient retrouvées dans le mix de plasmide parental ayant servi à la transfection des modèles cellulaires. Les barres noires représentent les fréquences auxquelles sont retrouvées différentes tailles de triplets répétés après réplication des cellules en cultures. Les tailles de triplets sont réparties en 3 catégories « moins de 79 répétitions » ; « 79 répétitions » ; « plus de 79 répétitions ». (Adapté de Cleary et al. 2002).

Il résulte de cette expérience qu'un éloignement de l'OR en amont et en aval des répétitions induit un biais vers les contractions. Cependant, dans cette expérience il est intéressant de constater qu'à une distance d'environs 100nt, si l'origine de réplication est en amont des CTG, aucune instabilité n'est détectée tandis qu'en aval, l'instabilité est biaisée vers les expansions. Ces résultats sont compatibles avec les observations de Liu *et al.* 2012

dans des cellules de patients. Ainsi, les expansions de répétitions CTG auraient préférentiellement lieu au cours de l'élongation débutant en aval des CTG. lorsque le brin porteur de CAG sert de brin modèle, c'est-à-dire lorsque le brin porteur de CTG est synthétisé. Or, les répétitions CTG forment des hairpins plus stables que les répétitions CAG. Ces hairpins vont provoquer un recul de l'ADN polymérase qui répliquera alors plusieurs fois les mêmes répétitions modèles (voir chapitre III.4.1).

III.3.4.4. EPIGENETIQUE ET SITES DE FIXATION DE CTCF

Le métabolisme de l'ADN est étroitement régulé, notamment par les marques épigénétiques. La marque épigénétique la plus étudiée est la méthylation, connue pour être impliquée dans la régulation normale ou pathologique (cancers) de l'expression génétique. La méthylation a pour la première fois été identifiée dans les séquences CpG où un groupe méthyl se rajoute directement sur les cytosines. Depuis, il est aussi reconnu que la méthylation intervient sur les histones et que la méthylation est conservée au fil des divisions cellulaires. La méthylation peut entrainer l'hétérochromatisation de l'ADN (1) par repliement et (2) par compétition : les sites méthylés ne peuvent plus fixer certains facteurs de transcription, ou d'autres facteurs régulateurs. L'expression de séquences à triplets répétés CGG et CCG (responsables du syndrome de l'X fragile et syndromes associés) peut être régulée par la méthylation *via* différents mécanismes de régulation (Sweatt *et al.* 2013):

- Diffusion de la méthylation depuis les séquences environnantes sur les séquences répétées méthylables telles les répétitions CGG;
- (2) Action de siRNA issus de la transcription bidirectionnelle pouvant induire le recrutement d'histone methyl-transférase ;
- (3) Perte de la capacité fixation de protéines isolant l'hétérochromatine de l'euchromatine telles que CTCF (CCCTC-binding factor): il en découle un phénomène d'hétéro-chromatisation aboutissant à une hyper-méthylation des séquences alentours.

Il a été démontré que le statut de méthylation influence l'instabilité des TNRs, notamment dans différents modèles de la maladie de Huntington, de la DM1, de l'ataxie de Friedreich, ou encore du syndrome de l'X fragile (voir (Robertson 2005, Pook 2012, Dion 2014, Nageshwaran *et al.* 2015) pour revues).

Par exemple, il est connu que l'instabilité intergénérationnelle du Syndrome de l'X Fragile est dû à des évènements prézygotiques, avec un effet du sexe du parent transmetteur : les allèles maternels transmettent des expansions, les allèles paternels des contractions. Or, il a été prouvé que les allèles paternels subissent une déméthylation rapide (Tamanini *et al.* 1997) avant que ne commencent la réplication des cellules du zygote ; contrairement aux allèles maternels qui eux conservent la méthylation de leurs CpG. De plus, il a été démontré dans un modèle de cellules COS1 porteuses d'épisomes à répétitions CGG que les délétions sont le résultat de la réplication, dont la dynamique est liée au statut de méthylation: si les régions riches en CpG sont méthylées, les auteurs ont observé une diminution de l'amplitude des variations de tailles des répétitions et l'inhibition des contractions au cours de la réplication dans les cellules (Nichol Edamura *et al.* 2005). L'ensemble de ces données suggère que les délétions paternelles de triplets CGG seraient le résultat de la déméthylation rapide des îlots CpG.

L'influence de la méthylation sur les triplets CAG impliqués dans la SCA1 a aussi été étudiée. Dans un modèle de cellules humaines de la SCA1, l'inhibition spécifique de *Dnmt1* (responsable du maintien de la méthylation des CpG) favorise l'instabilité somatique ; les auteurs ont également observé que l'inhibition de *Dnmt1*, dans un modèle murin de la SCA1, favorise l'instabilité des triplets CAG au cours des transmissions paternelles et maternelles avec un biais vers les contractions, ces contractions étant plus importantes dans les transmissions femelles. De plus, la fréquence de transmissions stables dans ces conditions était plus élevée dans le cas de transmissions mâles.(Dion *et al.* 2008) : méthylation et sexe du parent transmetteurs seraient donc co-acteurs de l'instabilité intergénérationnelle dans la SCA1.

Autre exemple, il a été identifié des sites de fixation du CTCF au voisinage de différents gènes responsables de pathologies à triplets répétés. Ce site est capable de fixer une protéine en doigt de zinc qui viendra moduler l'expression des gènes voisins, en modifiant la structure de la chromatine qui empêchera alors toute interaction des promoteurs avec d'éventuels « enhancers » ou « silencers ». Des études ont montré que CTCF ne se fixait plus à l'ADN au voisinage des grandes amplifications CTG dans le cas de la DM1 congénitale (Filippova et al. 2001, Cho et al. 2005, Filippova 2008, Yanovsky-Dagan et al. 2015). Cette observation est couplée avec le fait que la séquence cible de CTCF est hyper-méthylée en cas de CDM, empêchant ainsi sa fixation. La non-fixation de CTCF entrainerait à son tour la méthylation des séquences environnantes par diffusion. La même observation a été faite dans le voisinage des gènes FMR1, frataxine ou encore ATXN7 (Greene et al. 2007, Libby et al. 2008, Sun et al. 2018). Dans le cas de la SCA7, Libby et al. ont démontré que la tissus spécificité de l'instabilité somatique était liée au pourcentage de méthylation des séquences environnantes, avec une instabilité accrue en cas d'hyperméthylation invalidant la fixation du CTCF : la méthylation et le CTCF auraient donc une influence particulièrement importante sur l'instabilité somatique. Par ailleurs, plusieurs études dans des cellules de patients hESCs et dans un modèle murin DM1 ont démontré une corrélation positive entre la taille des répétitions de CTG et le taux de méthylation des séquences environnantes. Cependant, dans le cas de la forme congénitale,

cette corrélation est imparfaite (Brouwer *et al.* 2013, Yanovsky-Dagan *et al.* 2015, Barbé *et al.* 2017)

Enfin, une étude dans des cellules hESCs et CVS dérivées de patients DM1 et de contrôles non atteints, a démontré que les séquences environnantes des répétitions CTG n'étaient méthylées qu'en cas de répétitions pathologiques (Barbé *et al.* 2017). De plus, le taux de méthylation était d'autant plus élevé que l'âge d'apparition des symptômes était réduit. En effet, dans cette étude, les plus forts taux de méthylation étaient observés dans les cas de DM1 congénitale (CDM). Or, la forme congénitale est associée aux plus larges expansions de CTG connues dans la DM1, héritée par transmission maternelle. De plus, les auteurs ont observé une polarisation de la méthylation en fonction de la taille des répétitions : les patients CDM présentent majoritairement une hypermétylation en amont *et* en aval des répétitions, ce n'est qu'en aval des répétitions *ou* en amont. Les auteurs supposent que les spermatozoïdes avec une hyper méthylation en amont des répétitions ne survivent pas à l'invalidation de *SIX5* dont le promoteur est alors méthylé. Cette étude propose une explication à l'influence du parent transmetteur dans l'instabilité intergénérationnelle de la DM1 en lien avec la méthylation des séquences environnantes.

L'ensemble de ces données démontre l'impact de la méthylation sur l'instabilité intergénérationnelle et somatique de triplets de différentes natures, sur différents *loci*. D'autres études, moins nombreuses, ont démontré que d'autres marques épigénétiques impactaient l'instabilité comme l'acétylation des histones (voir (Evans-Galea *et al.* 2013, Usdin *et al.* 2015) pour revues).

A travers ce chapitre, nous avons pu illustrer l'importance de l'influence des triplets répétés en eux-mêmes (longueur des répétitions, pureté des répétitions, structures secondaires), mais aussi des séquences environnantes (méthylation, SNP) sur l'instabilité des TNRs. Le prochain chapitre sera lui dédié à la description de l'impact du métabolisme de l'ADN sur l'instabilité des triplets répétés, à savoir la réplication, la réparation et la transcription.

III.4. METABOLISMES DE L'ADN ET INSTABILITE

III.4.1. LA REPLICATION

La réplication de l'ADN est hautement liée à l'état de repliement de l'ADN, en effet, afin de pouvoir se poursuivre, toute une partie de la machinerie réplicative est dédiée à la résolution des structures secondaires. Comme nous l'avons vu précédemment, les répétitions anormales de triplets modifient grandement la structure de l'ADN, ce qui influencera donc la réplication. La réplication est découpée en trois grandes étapes : l'initiation, l'élongation et la terminaison. Nous aborderons dans ce chapitre le rôle de la réplication dans l'instabilité des TNRs à chacune de ces étapes.

III.4.1.1 INITIATION DE LA REPLICATION

Chez l'Homme, à chaque cycle cellulaire, ce sont entre 30 000-50 000 origines de réplication (ORs) qui sont activées, bien que le génome en contienne environs 5 fois plus : selon le cycle ou les cellules, ce ne seront pas toujours les mêmes origines qui seront sollicitées, on parlera alors de « flexibilité de réplicon » (Cayrou *et al.* 2011). Le motif exact des ORs humaines n'est pas connu, et il est suggéré que la structure de l'ADN a un rôle prédominant dans la reconnaissance des ORs. D'autant plus que les ORs humaines sont retrouvées dans des environnements riches en séquences CpG formant de nombreuses structures secondaires, majoritairement des structures G-quadruplexes, contrairement aux bactéries où les origines de réplications sont riches en AT (Leonard *et al.* 2013). Les origines de réplication sont reconnues par le complexe ORC (Origin Replication Complex) qui va recruter différents facteurs, notamment les hélicases, menant à l'ouverture de la double hélice par rupture des liaisons phospho-diesther, les protéines RPA seront ensuite recrutées afin d'éviter le repliement en double hélice de l'ADN. Cette structure particulière de l'ADN est appelée « Œil de Réplication ».

Comme nous l'avons vu dans le chapitre III.3.4.3., le positionnement de l'origine de la réplication va être déterminant pour l'instabilité, tant en fonction de sa distance que de son orientation vis-à-vis des répétitions pathologiques. Deux modèles d'influences des origines de réplications ont été proposés (voir (Cleary *et al.* 2005) pour revue et figure 13).

MODELE « ORIGIN-SWITCH »

Le modèle « origin switch » repose sur le fait qu'une fois le nombre seuil de répétitions dépassé, l'origine de réplication majoritairement utilisée peut être changée, impliquant une substitution du brin modèle par son brin complémentaire. Ce « switch » s'expliquerait par la reconnaissance et le « choix » de l'origine de réplication, qui dépendent fortement des

structures secondaires, or, celles-ci sont modifiées par les répétitions pathologiques. Différentes preuves en faveur de ce modèle ont émergé pour différentes maladies à triplets. Dans le cas de l'ataxie de Friedreich, une étude dans des cellules dérivées de patients a mis en évidence que le locus mutant n'était répliqué qu'avec des origines de réplication « dormantes » dans la population générale (Stevanoni et al. 2016). Pour le syndrome de l'X Fragile, un modèle de cellules hESC a démontré une absence de réplication par l'origine en amont des répétitions, contrairement aux d'individus contrôles (Gerhardt et al. 2014). Par ailleurs, il a été démontré dans un modèle murin de la DM1 porteur d'une copie du locus DMPK humain, que la seule origine de réplication active sur l'allèle mutant était celle située en aval des répétitions. Or, chez les patients DM1 (porteur d'un allèle mutant et d'un allèle normal) les 2 origines (en amont et en aval) de DMPK sont actives. Ce résultat suggère que chez les patients il pourrait il y avoir un changement de « choix » de l'origine de réplication à cause des répétitions CTG anormales : l'allèle saint serait donc répliqué par l'origine de réplication en amont (et peut-être en aval), tandis que le *locus* mutant ne serait répliqué qu'avec l'OR en aval (Cleary et al. 2010). Cependant, il est difficile chez l'Homme de distinguer l'activité des ORI de l'allèle normal par rapport au pathologique.

MODELE « ORIGIN-SHIFT »

Ce modèle suggère un décalage de quelques paires de bases de l'origine de réplication sur le *locus* mutant une fois le nombre de répétitions pathologique seuil atteint : l'initiation de la réplication est décalée, mais le sens de progression qui en découlera reste le même, donc le brin modèle restera le même. Cependant, comme nous l'avons décrit dans le chapitre III.3.4.3., un tel changement peut avoir des conséquences drastiques sur l'instabilité et le biais de l'instabilité vers les expansions/contractions. Ce modèle impacte l'instabilité par la délocalisation de la zone d'initiation d'Okazaki (voir chapitre suivant) par rapport aux répétitions pathologiques. Il a ainsi été démontré que les « shifts » des origines de réplication dans l'environnement proche des TNRs peuvent influencer l'instabilité en modifiant sa nature (biais vers les expansions ou les contractions), ou en modulant l'amplitude de l'instabilité (stabilisation ou déstabilisation des TNRs) dans le cas de répétitions impliquées dans la DM1, le FRAXA, et la DM2 (voir (Cleary *et al.* 2005) pour revue).



Figure 13: Schématisation des modèles "origin switch" et "origin shift".

Ori: origine de réplication; cercle vert: ori active; cercle gris: ori inactive; flèche: sens de progression de la réplication.

LE MODELE FORK SHIFT

Dans ce modèle, les auteurs proposent que la fourche de réplication soit modulée par des éléments en *cis* (représentés par l'éclair sur la figure 14) qui mène à un décalage de la zone d'initiation des fragments d'Okazaki (ZIO) (Cleary *et al.* 2005). Il peut s'agir de modifications épigénétiques, de mutations, d'autres processus cellulaires entrant en conflit... Dans ce modèle, les brins continu et indirect restent les mêmes et le sens de la réplication est conservé. Seul le « choix » de la zone d'initiation d'Okazaki sur les brins parentaux changera. En effet, les ZIO ne correspondent qu'à certaines natures de séquences (séquences « incluses »), plus ou moins fréquemment utilisée pour la réplication. Un bouleversement menant à l'incapacité d'utiliser une ZIO aboutira à une initiation décalée vers la prochaine ZIO compatible. Ainsi la position des séquences dans la fourche de réplication (amorce ARN, brin synthétisé, jonction entre deux fragments d'Okazaki...) sera changée.

L'instabilité découlant d'un « Fork Shift » dépendra de la nature des triplets, de la longueur de la répétition et de la proportion de triplets répétés comprise dans la ZIO. Ces facteurs modulent en effet la formation de structures secondaires, l'interaction de l'ADN avec des protéines de réparation liées à la réplication, et, de manière générale, la dynamique de la réplication.



Figure 14: Schématisation du modèle "Fork Shift".

La dynamique de la fourche de réplication est modifiée par des éléments en cis de la séquence répliquée, menant à un décalage de la zone d'initiation des fragments d'Okazaki. Ori: origine de réplication; cercle vert: ori active; cercle gris: ori inactive; ZIO : Zone d'Initiation des fragments d'Okazaki ; éclair : évènement perturbateur en cis menant au décalage de la ZOI.

Si l'initiation de la réplication conditionne l'instabilité en géo-localisant la réplication, elle n'en est pas directement responsable. En effet, une étude dans des cellules primaires de patients DM1 a démontré que si l'initiation de la réplication était inhibée (par mimosine) l'instabilité des triplets n'en était pas changée (Yang *et al.* 2003). En revanche, si l'élongation par l'avancement de la fourche de réplication est inhibée, cela favorise l'instabilité de l'allèle mutant.

III.4.1.2. L'ELONGATION : PROGRESSION DE LA FOURCHE DE REPLICATION

Une fois l'ouverture de la double hélice stabilisée, l'ADN polymérase progresse le long de l'ADN à répliquer : on parlera de progression de la fourche de réplication. Cette progression se fait selon deux modes en fonction du sens de la réplication du brin par rapport à la fourche de réplication (voir figure 15). Le brin parental direct est synthétisé de façon continue, tandis que le brin parental indirect est produit par fragments (fragment d'Okazaki) de 100-200pb chez les eucaryotes. La production du brin retardé débute par la synthèse d'amorces ARN par la Primase. Ces amorces forment des complexes ADN/ARN qui serviront de point de départ à l'ADN polymérase pour la synthèse de fragment d'Okazaki de 5' en 3'. Les amorces ARN sont ensuite détruites entre autres par l'ARNase H et FEN1. Enfin, les fragments d'Okazaki sont reliés les uns aux autres par la Ligase I. L'élongation de la réplication est un point critique pour l'instabilité du génome en général, car elle génère de l'ADN simple brin potentiellement reconnu par des mécanismes de réparation, elle fragilise certaines zones du génome par super enroulement pouvant mener à des cassures simples et doubles brins, et elle peut être sujette à des erreurs.



Figure 15: Fourche de Réplication active.

Le brin retardé a été replié par les protéines RPA pour permettre l'accès à l'ADN polymérase sur le brin retardé, et permet le rapprochement des fragments d'Okazaki, ce qui facilite leur suture. Sur ce schéma, l'ADN polymérase est sur le point de se décrocher du brin d'ADN (fragment d'Okazaki) qu'elle vient de synthétiser, avant passer à la prochaine amorce ARN. Figure traduite d'après Walter, Bruce Alberts Alexander Johnson Julian Lewis Martin Raff Keith Roberts Peter. Molecular Biology of the Cell. 5th ed.; p276. Edited by Garland Science. 2008. Les répétitions de triplets représentent des zones particulièrement à risque au cours de l'élongation de la réplication et demandent le concours d'hélicases spécifiques pour pouvoir être répliquées. Les hélicases FANCJ sont par exemples dédiées à la résolution de structures G4s. Chez l'Homme, l'inhibition de FANCJ entraîne une instabilité globale dans les zones de triplets CTG/CAG répétés du génome (Barthelemy *et al.* 2016). Ou encore, un modèle cellulaire humain de la DM1 a démontré que l'hélicase RTEL1, impliquée dans l'ouverture de structures secondaires, et régulant négativement la recombinaison homologue, aide à empêcher les expansions de triplets CTG (Frizzell *et al.* 2014). Nous décrirons plus en détails dans les paragraphes suivant l'interdépendance entre la dynamique de la fourche de réplication et les triplets répétés.

La dynamique de l'élongation est finement régulée, et est cruciale pour préserver l'intégrité du génome en général. Par exemple, il a été démontré que si la fourche de réplication est accélérée, l'instabilité du génome s'en retrouve augmentée (Maya-Mendoza *et al.* 2018). Quant à l'arrêt de la fourche c'est un élément de la dynamique de la réplication qui se produit naturellement dans les cellules (malades ou normales). Cependant, comme nous le décrirons dans les paragraphes suivants, les répétitions de triplets pathologiques peuvent mettre à mal ce phénomène.

ARRET DE LA FOURCHE DE REPLICATION

Les pauses de la réplication interviennent naturellement en fonction des étapes du cycle cellulaire et sont un élément garant de l'intégrité du génome. Cependant, l'arrêt de la réplication sur l'un des deux brins peut provoquer une désynchronisation entre la réplication du brin continu et du brin discontinu. Ce qui augmente anormalement la quantité d'ADN simple brin dans la fourche de réplication ce qui favorise la formation de structures secondaires aberrantes et le recrutement de mécanismes de réparation pouvant aboutir à de l'instabilité (Mirkin *et al.* 2007). Si les pauses sont reconnues comme anormales (trop longues, n'intervenant pas au bon moment dans le cycle cellulaire...) la cellule y voit un message d'erreur qu'elle tentera de résoudre par différents moyens : cassure, recombinaison, redémarrage de la fourche.

La stabilité de la fourche de réplication est indispensable à l'intégrité du génome, notamment dans l'instabilité des triplets répétés. En effet, l'inhibition de stabilisateurs de la fourche Timeless, Clapsin et Tipin, mène à l'effondrement de la fourche de réplication, et augmente l'instabilité des triplets CTG dans un modèle de cellules Hela DM1 (Liu *et al.* 2010). Or, des modèles de levures, bactériens et de cellules humaines ont prouvé que les répétitions CGG, CTG et GAA provoquent des pauses de la fourche de réplication (Samadashwily *et al.* 1997, Pelletier *et al.* 2003, Krasilnikova *et al.* 2004, Follonier *et al.* 2013). Plus précisément,

un modèle de cellules humaines a mis en évidence que les pauses de la réplication au niveau des répétitions GAA et CTG étaient dépendantes des structures secondaires formées par les TNRs (Voineagu *et al.* 2008, Gerhardt *et al.* 2016). Cet effet est d'autant plus marqué en combinaison avec le stress réplicatif qui peut imposer la formation de structures secondaires non canoniques sur les répétitions (Liu *et al.* 2010).

Les triplets répétés influencent la dynamique de la fourche de réplication (arrêt, déplacement des ZOI), et peuvent mener, du fait de leur monotonie, à des dérapages de la polymérase, ce qui favorise l'insertion ou la délétion de nucléotides, donc l'instabilité des TNRs. Cependant, le redémarrage de la fourche de réplication joue également un rôle prépondérant dans l'instabilité des triplets.

<u>REDEMARRAGE ET DÉRAPAGE DE LA FOURCHE DE REPLICATION</u>

Après une pause anormale de la réplication, plusieurs scénarii de réparation postréplication (PRR) sont possibles. Tout d'abord les mécanismes dépendants d'un « template switch » : la synthèse translétionnelle (TLS), et la recombinaison homologue. Ces mécanismes reposent sur le dépassement de l'élément perturbateur (ex. : hairpin) par la fourche de réplication. D'autre part, le redémarrage de la réplication peut passer par le recul de la fourche, et la re-réplication de la zone entre la zone d'initialisation et l'élément perturbateur qui aura été corrigé (voir (Mirkin *et al.* 2007, Saugar *et al.* 2014, Marians 2018) et résumé figure 16).

La réplication par synthèse translétionnelle (TLS) repose sur le recrutement d'ADN polymérases de type Y, qui sont capables de passer outre une base endommagée ou un mésappariement. Ces polymérases ne disposent pas d'activité exonucléasique, leur recrutement est donc source d'erreurs réplicatives. L'implication de la TLS dans l'instabilité des triplets répétés est sujet à débat. En effet, il a longtemps été admis que ce système n'influençaient pas l'instabilité des triplets répétés, mais certaines études ont plus récemment mis en avant que, même si limité, ce système aurait une influence sur les répétitions GAA et CAG et favoriseraient la formation d'hairpins courtes qui mèneraient à des expansions de quelques nucléotides (voir (Usdin et al. 2015) pour revue et (Chan et al. 2013)). La recombinaison homologue liée à la réplication intervient après la détection de l'état simple brin anormal de l'ADN néoformé. Cette situation peut être due à l'arrêt de la réplication sur l'un des brins d'ADN ou faire suite à une cassure sur l'un des brins lors d'un stress réplicatif (surenroulement, encombrement...). La machinerie de réplication utilise alors le brin opposé comme modèle afin de synthétiser la séquence manquante. Cependant, ce mécanisme repose en partie sur le bon alignement des brins d'ADN complémentaires, et la monotonie des répétitions peut mener à un décalage d'alignement, ce qui aboutit à des expansions ou des

contractions. De nombreux travaux dans des modèles de levures ont démontré l'importance de ce phénomène dans l'instabilité des triplets répétés (voir (Polleys *et al.* 2017) pour revue et chapitre III.4.2 où nous décrirons plus en détail ce mécanisme de réparation). **Le recul de la fourche de réplication** se fait après régression et dégradation des brins néo-formés qui se retrouvaient à l'intérieur de la fourche (voir figure 16) par le mécanisme BIR (Break Induced Replication). Ce mécanisme mène à la cassure double brin de l'ADN néo-formé ce qui revient à un « recul » de la fourche de réplication. La fourche tente alors de poursuivre la réplication. Cela peut passer par l'élimination de « l'élément perturbateur » par des mécanismes de réparation, ou par dépassement de l'élément perturbateur (par recombinaison homologue). Ce mécanisme est notamment connu pour permettre la réplication des télomères et induire de larges expansions dans les tumeurs (voir (Leffak 2017) pour revue). La plupart des données sur ce mécanisme viennent d'études sur la levure, où il a été démontré que les protéines du BIR étaient nécessaires pour la génération de larges expansions, notamment pour le cas des triplets CAG (Kim *et al.* 2017). Un modèle ce cellule humaine a par ailleurs confirmé l'implication de ce mécanisme dans les larges expansions CGG (Kononenko *et al.* 2018).



Figure 16: Arrêt de la fourche de réplication et différentes voies de dépassement de lésion.

Ici la lésion est représentée par une hairpin. Si le redémarrage de la réplication se fait en amont de la lésion, lieu de décrochage de l'ADN polymérase, alors la réplication aboutira à une expansion. Si la réplication reprend après la lésion, la réplication aboutira à une expansion ou une contraction. A l'exception du redémarrage par TLS où la polymérase recrutée est beaucoup moins fidèle et insèrera des insertions/délétions indépendamment de la lésion. Les brins parentaux sont en bleu, le brin néoformé direct est en orange, le retardé en vert. La reprise de la polymérisation est symbolisée par les flèches violettes. Figure traduite et adaptée d'après Marians et al. 2018.
<u>RETRAIT DES AMORCES ARN ET LIGATION DES FRAGMENTS D'OKAZAKI</u>

Une fois les fragments d'Okazaki du brin retardé synthétisés, différents acteurs sont sollicités afin de les relier entre eux pour former un brin complet, ininterrompu. Pour cela, les complexes ADN/ARN servant d'amorces doivent être éliminés (notamment via l'endonucléase FEN1), et les fragments d'Okazaki religués entre-eux (via la ligase LIG-1).

FEN1 (Flap Endonucléase 1) est une nucléase structure-spécifique, spécialisée dans l'élimination des structures ADN/ARN (flaps) générées lors de la maturation des fragments d'Okazaki, elle intervient aussi dans le mécanisme de réparation BER, NHEJ et la recombinaison homologue. Si FEN1 ne peut pas accéder au brin néo-formé à cause de structures secondaires anormales, l'ADN des flaps reste en place et est intégré au néo-brin retardé : cela mène donc à des expansions. Il a été démontré in vitro dans des modèles de levures et de cellules humaines, que l'activité nucléase de FEN1 était bloquée par les structures secondaires formées par les répétitions anormales de triplets, et qu'il en résultait des expansions dues à la réplication (Spiro et al. 1999, Henricksen et al. 2000, Lee et al. 2002, Lai et al. 2013). L'ablation totale de FEN1 est létale, c'est pourquoi les seules études in vivo dans des modèles murins repose sur l'analyse de l'impact de seulement une diminution de l'expression de FEN1. Cependant, la seule diminution de FEN1 dans un modèle murin de la maladie de Huntington est suffisante pour augmenter la fréquence des expansions CAG (Spiro et al. 2003), mais elle n'impacte pas l'instabilité des triplets CTG dans un modèle murin de la DM1 (van den Broek et al. 2006). Ce qui laisserait supposer que selon la nature et/ou la localisation génique des triplets, FEN-1 seule n'est pas forcément capable d'en impacter l'instabilité. De même, lorsque la protéine DNA Ligase 1 (orthologue de Cdc9 chez la levure) est invalidée dans des modèles de levures, cela favorise l'instabilité des triplets CAG vers les expansion (Refsland et al. 2005, Richard et al. 2008).

III.4.1.3. TERMINAISON DE LA REPLICATION

Chez les eucaryotes, la terminaison de la réplication se fait lorsque 2 fourches progressant en sens inverse sur le même brin d'ADN se rencontre et fusionnent ensembles. La jointure entre 2 fourches de réplication se fait généralement de façon stochastique sur le génome, mais il existe aussi des zones dites RFBs (Replication Fork Barriers) de terminaisons de la réplication qui ont été identifiées notamment chez l'Homme et la levure. Ces zones permettent de stopper une fourche, jusqu'à ce que la seconde l'y rejoigne et jouent le rôle de terminateurs, évitant la collision des fourches en faveur d'une fusion contrôlée (Dewar *et al.* 2017, Gold *et al.* 2017). Cependant, des études chez la bactérie et la levure ont démontré que ces zones étaient des points chauds d'instabilités. En effet, les fourches y sont stoppées, ce qui peut mener à tous les mécanismes d'instabilité décrits dans les paragraphes suivants. Enfin, les RFBs sont aussi un point de régulation des avancées potentiellement conflictuelles

de la fourche de réplication et de la transcription. Une collision entre la réplication et la transcription peut effectivement être vectrice d'instabilité car cela recrute des mécanismes de réparation qui peuvent induire de l'instabilité (voir chapitre III.4.2). Les mécanismes déterminant les propriétés des RFBs sont peu connus à ce jour. Cependant, un modèle de levure *S. pombe* porteur de 70 CAG à différentes distances d'une séquence RFBs n'a pas démontré de lien entre ces séquences et l'instabilité des triplets (Gold *et al.* 2017).

À travers ce chapitre, nous avons pu décrire le lien étroit entre la réplication et l'instabilité. Comme nous l'avons vu, l'instabilité médiée par la réplication était étroitement liée aux structures secondaires et aux séquences environnantes des répétitions, mais également aux mécanismes de réparations. Nous nous proposons dans la suite de ce manuscrit de décrire plus en détail l'impact de la réparation sur l'instabilité des triplets.

III.4.2. LA REPARATION

III.4.2.1 LES MECANISMES DE REPARATION PAR EXCISION DE BASES

L'instabilité des triplets répétés est détectée dans les cellules mitotiques et non mitotiques, de plus, différentes études dans des modèles murins ont démontré que le niveau d'instabilité des triplets n'était pas corrélé aux l'activités mitotique et transcriptionnelle (Lia *et al.* 1998, Gonitel *et al.* 2008, Gomes-Pereira *et al.* 2014). Ces observations ont suggéré que les mécanismes de réparation pouvaient être impliqués dans l'instabilité des TNRs. Depuis, de nombreuses protéines des différents mécanismes de réparation ont effectivement été reliées à l'instabilité. Dans ce chapitre, nous passerons en revue les différents mécanismes de réparation et leur influence sur l'instabilité des TNRs.

• <u>LE BER</u>

Le BER (ou « Système de Réparation par Excision d'une Base) traite différents types de lésions comme les bases désaminées, alkylées ou oxydées. Ce système repose sur l'hydrolyse spontanée de la liaison glycosidique de la base endommagée, ou sur l'activité de différentes ADN glycosylases, chacune spécialisée dans le traitement d'un de ces dommages. Ces glycosylases cheminent le long de l'ADN et s'arrêtent lorsqu'elles détectent une lésion. À ce moment-là, la glycosylase retire la base incriminée, laissant un désoxyribose abasique reconnu par l'endonucléase APE1 (Endonucléase Apurinique) ou la lyase AP, qui génèrent une délétion d'une base. S'en suit alors le remplacement de la base selon 2 mécanismes différents :

-le système « SN-BER » (Single-nucléotide BER), impliquant l'ADN polymérase ß qui insère le nucléotide manquant, les liaisons se feront grâce au facteur XRCC1 et l'ADN ligase I (LIG1) ou l'ADN ligase III ;

- **le système LP-BER** (Long-Patch BER) qui repose sur la même mécanistique que la réplication du brin indirect, avec le recrutement de l'ADN polymérase ß qui va synthétiser au minimum 2 nucléotides. Cette synthèse va générer un court flap 5' qui sera éliminé par Fan1, puis la LIG1 relie les bases synthétisées au brin d'ADN endommagé. Ce système permet de réparer au minimum 2 nucléotides.

Le rôle du BER dans l'instabilité des triplets a été suggéré par l'augmentation de l'instabilité dans des fibroblastes HD et chez des souris pré-mutées pour FXS exposés à l'eau oxygénée au bromure de potassium respectivement, agents favorisant les dommages oxydatifs sur les nucléotides (Kovtun *et al.* 2007, Entezam *et al.* 2010). Depuis, différentes protéines effectrices du BER ont été identifiées comme impliquées dans d'instabilité des TNRs (voir tableau 6).

Protéine	Activité initiale de la protéine	Contexte génomique	Conséquence de l'invalidation de la protéines sur l'instabilité	Référence
LIG1	Suture de fragments d'ADN au cours de la réplication et du SN-BER	Souris DM1 DM300/ <i>lig1</i> hypomorphique homozygote	Diminue la fréquence des expansions, augmente la fréquence des contractions dans la lignée germinales maternelles des CTG	(Tomé <i>et al.</i> 2011)
		Souris FXS KI C57BL/6 <i>lig1</i> hypomorphique homozygote	Aucun impact sur l'instabilité intergénérationnelle et somatique des CGG	(Entezam <i>et al.</i> 2010)
OGG1	Glycosylase spécifique du retrait d'erreur 8- oxoG	Souris HD (Q150/wt) /ogg1(-/-) et (Q150/Q150) /ogg(-/-)	Perte des expansions somatiques	(Budworth <i>et al.</i> 2015)
		Souris Svj129 avec de longues répétitions CAG sur 30 <i>loci</i> mais non pathologique/ KO ogg1	Augmentation de l'instabilité somatique sur <i>locus tbp</i> avec faibles variations de tailles de répétitions	(Sanchez- Contreras <i>et al.</i> 2017)
		Souris HD R6/1 KO ogg1	Élimination de la corrélation âge/ expansion des CAG en diminuant l'instabilité somatique	(Kovtun <i>et al.</i> 2007)
NEIL1	Endonucléase	Souris HD R6/1 KO <i>neil1</i>	Réduction de l'amplitude des expansion intergénérationnelles, augmentation de la fréquence des expansions intergénérationnelles et somatiques des CAG	(Mollersen <i>et al.</i> 2012)
		Souris HD R6/1 KO neil1	Diminution la fréquence des expansions dans l'instabilité somatique des CAG	(Kovtun <i>et al.</i> 2007)
FEN1	Endonucléase spécifique des structures flaps 5'	Souris HD KO R6/1 fen <i>1</i> hétérozygote	Diminue la fréquence des contractions intergénérationnelles en faveur des expansions dans les transmissions paternelles des CAG	(Spiro <i>et al.</i> 2003)
		Souris FXS KI C57/BL6 KO fen1 hétérozygote	Aucun effet sur l'instabilité intergénérationnelle et somatique des CGG	(Entezam <i>et al.</i> 2010)
		Souris DM1KI DM1 (C57BL/6) / KO <i>fen1</i> hétérozygote	Aucun effet sur l'instabilité somatique des CTG	(van den Broek <i>et al.</i> 2006)

Tableau 6: Protéines du BER impliquées dans l'instabilité des triplets dans des modèles murins.

Plusieurs modèles expliquant l'implication du BER dans les expansions des triplets ont été proposés *via* le LP-BER (voir (Goula *et al.* 2013) pour revue). Le premier modèle fait suite à l'observation de pauses dans la réparation de lésions 8-oxoG au voisinage de répétitions CAG en lien avec la formation d'hairpins dans un modèle de cellules murines (Liu *et al.* 2009). Dans ce modèle, l'équilibre entre l'activité de l'ADN polymérase ß et FEN1 est corrompu. Les expansions résultent de l'incapacité de FEN1 à cliver l'intégralité des flaps générés par la polymérase : les flaps se retrouvent alors intégrés au brin réparé (Goula *et al.* 2012). Le second modèle se base sur la capacité de l'ADN polymérase ß de rajouter des nucléotides à l'extrémité 3' des hairpin. Ces nucléotides servent de point de départ pour l'ADN polymérase δ (Pol δ) qui va poursuivre la synthèse de façon continue, incorporant ainsi l'hairpin au brin

réparé: ce rajout de nucléotide va aboutir à une expansion (Chan *et al.* 2013). Enfin, la formation d'hairpin est aussi déterminante pour l'activité d'OGG1.

Le BER est aussi impliqué dans les contractions de triplets. En effet, il a été observé dans des souris modèles de la HD, et des fibroblastes patients HD, que le stress oxydatif, favorisait les contractions. Plusieurs mécanismes pour tenter d'expliquer ce phénomène ont été proposés : la localisation de la base excisée pourrait influencer la balance expansion/contraction, ou bien l'ADN polymérase ß passerait outre l'hairpin repérée (voir (Goula *et al.* 2013, Usdin *et al.* 2015) pour revue et (Goula *et al.* 2013)). Une étude a démontré que le BER était vecteur de contraction dans un modèle *in vitro via OGG1*. L'étude se basait sur des oligonucléotides porteurs de 7 à 20 répétitions CAG formant des hairpins et soumis à des stress oxydatifs. Les guanines des boucles des hairpins étaient alors sous formes 8-oxoG. Les guanines ainsi transformées sont reconnues et retirées par OGG1, laissant un site abasique qui sera clivé par APE1. Ce clivage casse la boucle de l'hairpin qui est alors sous forme de deux flaps intermédiaires, clivés par FEN1 et d'autres endonucléases. Ce cheminement aboutit donc *in fine* à l'élimination de l'hairpin, ce qui provoque des contractions de triplets (Xu *et al.* 2014).

• <u>LE NER</u>

Le NER (ou « Système de Réparation par Excision de Nucléotides ») est un mécanisme de réparation se basant sur la détection de distorsion de la double hélice d'ADN plutôt que des modifications de bases. Ces modifications imposantes peuvent être par exemple le résultat de traitement par rayons ultra-violet (causant des dimères de thymine). Les lésions peuvent être reconnues par deux systèmes :

-le GGR (Global Genome Repair) qui répare des séquences sur l'ensemble du génome. Les lésions sont reconnues par le complexe XPC-hHR23B. Cependant, différentes études chez la souris et dans des cellules humaines ont démontré que ce système n'influence que peu l'instabilité des triplets (Lin *et al.* 2006, Dragileva *et al.* 2009)

-le TCR (Transcription Coupled Repair, ou TCNER) est spécifique des gènes transcrits. La lésion est reconnue par l'arrêt de l'ARN polymérase qui recrute alors les facteurs CSA et CSB (Cockayne Syndrome A et B) qui décrochent l'ARN polymérase de l'ADN. Les hélicases XPD et XPB maintiennent alors la double hélice ouverte et permettent de la rendre accessible aux mécanismes de réparation.

Dès la lésion identifiée, la séquence d'ADN incriminée (20-30 nucléotides) est éliminée par les protéines XPF-ERCC1 par clivage en 5' et en 3'. La réparation est ensuite assurée différentes ADN polymérases alliées à LIG1 ou LIG3/XRCC1.

L'implication du NER dans l'instabilité a d'abord été identifiée chez *E. Coli*, Cependant, les résultats étaient relativement contradictoires, par exemple, l'invalidation de la protéine de reconnaissance des lésion augmentait ou stabilisait les répétitions selon l'étude considérée (Parniewski *et al.* 1999, Oussatcheva *et al.* 2001, Szwarocka *et al.* 2007). Chez les eucaryotes, il est proposé que le NER intervienne dans l'instabilité des triplets répétés au moment de l'excision de la séquence lésée. En effet, l'invalidation de CSB diminue l'instabilité somatique ou la fréquence des expansions intergénérationnelles dans des modèles murins de FXS et HD respectivement (Kovtun *et al.* 2011). Ou encore, l'invalidation par siARN d'XPA réduit la fréquence des contractions dans un modèle de cellules humaines porteur de 92 CAG (Lin *et al.* 2012) et l'invalidation d'XPG, ERCC1 et CSB diminuent l'instabilité dans un modèle de cellule humaine à contraction de CAG (Lin *et al.* 2007). Ces résultats sont en cohérence avec la capacité d'XPA de fixer les hairpins (Lin *et al.* 2012) et d' XPG de les éliminer *in vitro (Hou et al.* 2011).

III.4.2.2. LES MECANISMES DE REPARATION DE CASSURES DOUBLE-BRINS PAR RECOMBINAISON

Les cassures doubles brins sont issues de (1) la formation de structures secondaires et les cassures mécaniques qu'elles provoquent au cours de la mitose ; (2) les défauts de réplication (glissement de brins néo-synthétisés, mauvais traitement des fragments d'Okazaki, résolution de l'arrêt de la fourche de réplication par recul...), (3) les mécanismes de réparation et (4) la recombinaison. Les cassures double brin (DSB) peuvent survenir au voisinage des triplets ou directement sur les triplets répétés.

En effet, il a été démontré dans des cultures de cellules humaines que les répétitions CGG/CCG dans le gène FMR1 sont sources de translocation à l'issue de cassures doubles brins (Sutherland *et al.* 1998, Yudkin *et al.* 2014). Les cassures double brin sont reconnues par les complexes ATM (*ataxia Telangectasia* Mutated) et ATR (AT- related protein) qui recruteront alors un des systèmes de réparation suivant : réparation homologue, non homologues et réparation des extrémités protubérantes.

LA RECOMBINAISON HOMOLOGUE

Dans le cas des recombinaisons homologues (HR), les DSB sont fixées chez l'Homme par le complexe MRN (composé des facteurs Mre11, Rad50 et Nbs1) qui permettra l'exposition des extrémité 3' simple brin. La protéine RPA se placera (grâce aux protéines Rad51, Rad52, Rad54) au niveau de la cassure afin de la protéger des nucléases et se déplacera en même temps que les acteurs de la réparation. La recombinaison à proprement parler dépend des facteurs Brca1 et Brca2 interagissant avec la recombinase Rad51. Rad51 reconnait les séquences homologues sur la chromatide sœur et permet l'invasion de brin. Une fois les chromatides sœurs alignée, la synthèse pour réparer la délétion peut se faire par homologie. La structure transitoire qui se forme alors est appelée structure de Holliday et sera éliminée par restauration de chacun des brins par des enzymes résolvases en fonction de la configuration de la jonction Holliday (voir (Wyatt *et al.* 2014) pour revue et figure 17).



Figure 17: Exemples de traitement des cassures doubles brin par la résolution des jonctions d'Holliday.

En vert : les différents acteurs des différentes étapes ; En bleu : la chromatide subissant la cassure double brin ; en orange : sa chromatide sœur ; en pointillé : les brins néosynthétisés au cours de la réparation. Adapté d'après ??

Un modèle bactérien a permis de mettre en évidence que les longues répétitions de CTG/CAG sont des points chauds de recombinaison homologue inter-moléculaire (Napierala *et al.* 2002, Pluciennik *et al.* 2002). Dans un modèle de levure porteur de répétitions CAG, après induction de DSB, il a été mis en évidence que le système MRX (équivalent du MRN humain) est impliqué dans la réparation des triplets, induisant majoritairement des expansions (Richard *et al.* 2000). De plus, ce modèle a démontré qu'en sur-exprimant *MRE11* ou *RAD50*, la longueur des expansions était plus grande. Par ailleurs, l'inhibition de *Rad52* dans un modèle murin de la DM1 porteur de 300 CTG diminue légèrement la taille moyenne des expansions (Savouret *et al.* 2003), tandis que, la même invalidation dans la levure, favorise l'instabilité (Sundararajan *et al.* 2010). Dans un modèle de cellule CHO DM1 (Chinese Hamster Ovary), il a été démontré que la recombinaison homologue induite favorisait cette fois-ci les larges contractions de répétitions CTG (Meservy *et al.* 2003).

La variabilité dans les résultats obtenus dans les différents modèles ainsi que le faible pourcentage de recombinaison détectée chez les patients DM1 aux *loci DMPK* et *FMR1* mutants laissent supposer que la recombinaison homologue n'influence que très peu l'instabilité des triplets (Fu *et al.* 1991).

LA RECOMBINAISON NON-HOMOLOGUE (NHEJ)

Contrairement au mécanisme de réparation homologue, le NHEJ n'utilise pas de modèle pour réparer les lésions : la réparation est faite en reliant directement les extrémités de la cassure *via* le complexe LIG4/XRCC4. Ce système mène ainsi intrinsèquement à de l'instabilité par insertion, délétion ou translocation et parfois perte d'hétérozygotie.

Un modèle de levure à répétition CAG délété pour *Dnl4* (équivalent de *Lig4* impliquée dans le NHEJ) présentait une fréquence de contraction significativement augmentée, ce qui laisse supposer que le NHEJ est lié à l'instabilité (Sundararajan *et al.* 2010). En cas de délétion de protéines nécessaires à la recombinaison homologue et non homologue, il est toujours observé des contractions chez la levure qui est la conséquence de l'action d'un autre mécanisme de réparation : la juxtaposition des extrémités (Sundararajan *et al.* 2010).

Le MMEJ (Microhomology Mediated End Joinning) est une alternative du NHEJ qui a besoin de micro-homologies pour réaligner les brins avant de combler la délétion avec l'ADN polymérase ß. La reconstitution *in vitro* de ce mécanisme a démontré que le MMEJ favorise les expansions de CAG (Crespan *et al.* 2012).

LA JUXTAPOSITION DES EXTREMITES ET RESECTION DES BRINS (SSA)

Ce mécanisme de réparation implique que la cassure double brin se soit faite avec des bouts francs, au milieu de séquences répétées homologues, en orientation directe. Les bouts francs vont être dégradés de 5' en 3', générant des extrémités sortantes 3'OH simples brins. Cette dégradation se poursuit jusqu'à ce que soient révélées des micro-homologies de part et d'autre de la cassure par les protéines MSH2 et MSH3. Les brins sont alors reséqués entre eux par alignement *via* RAD52 ou RAD51, ce qui génère des queues sortantes qui seront éliminées par les nucléases RAD1-RAD10. *In fine*, ce mécanisme mène donc à des délétions.

Dans un modèle de levure à répétitions CTG, il a été mis en évidence qu'en cas d'induction de cassures double brin, 67% des évènements de réparation étaient dus au mécanisme SSA, aboutissant ainsi à des contractions (Richard *et al.* 1999).

Les mécanismes NHEJ, MMEJ et SSA semblent prédominants lors de cassures double-brin induites et sont de potentiels alliés thérapeutiques. En effet, lors d'expériences en levures et cellules humaines utilisant des TALENs pour produire des DSB au sein de répétitions CAG/CTG, il a été observé une hausse de la fréquence des contractions (Mittelman *et al.* 2009, Richard *et al.* 2014).

III.4.2.3. LE SYSTEME DE REPARATION DES MESAPPARIEMENTS DE BASES (MMR)

Le système de réparation des mésappariements de base (MMR) consiste à réparer 1 ou plusieurs mésappariements (délétion, insertion, substitution). Ce système est indispensable à l'intégrité du génome : il corrige les erreurs de réplication, intervient dans la ligation des fragments d'Okazaki, limite les réarrangements chromosomiques et intervient dans différentes voies de réponses envers certains dommages de l'ADN (apoptose par exemple). Preuve de son indispensabilité, ce mécanisme est hautement conservé des procaryotes jusqu'aux eucaryotes supérieurs. Chez les eucaryotes, le MMR peut être recruté par d'autres mécanismes de réparation (comme le SSA), ou reconnait directement les lésions de l'ADN *via* des hétéro-dimères spécialisés : MSH1 intervient dans le maintien de l'ADN mitochondrial ; MSH2/MSH3 reconnaissent les boucles de mésappariement de taille importante ; MSH2/MSH6 reconnaitra les petites boucles et les substitutions ; tandis MSH4/MSH5 est spécialisé dans les évènements méiotiques.

Une fois l'erreur fixée, le MMR va recruter différents facteurs (MLH1, MLH2, MLH3, PMS1 et PMS2, PCNA) qui vont avoir un rôle stabilisateur dans le recrutement d'enzymes réparatrices en fonction du type de dommage reconnu, comme des exonucléases, protéines clamps, ADN polymérase, FEN1... (voir (Kolodner and Marsischky 1999, Norris 2015) pour revues, et figure 18).



Figure 18: Illustration du fonctionnement du MMR.

Le MMR est étroitement lié à l'instabilité des séquences microsatellites. Si le MMR est inhibé dans les cellules, celles-ci accumulent de nombreuses mutations et l'instabilité des séquences microsatellites est favorisée (Chen *et al.* 2005, Jiricny 2006). De plus, des mutations sur différents acteurs du MMR ont été reliées à des cas de cancers colorectaux familiaux HNPCC (Hereditary Non Polyposis Colorectal Cancer), où les cellules tumorales témoignent d'une grande instabilité des séquences microsatellites (Lynch 1999).

III.4.2.3.1. PREUVES DE L'IMPLICATION DU MMR DANS L'INSTABILITÉ DES TNRS CHEZ LES MAMMIFÈRES

Les effets du MMR sur l'instabilité des triplets répétés ont été le sujet de nombreuses publications dans de nombreux modèles eucaryotes et procaryotes. Cependant, l'analyse de modèles procaryotes et de certains eucaryotes (levures et drosophiles) n'ont pas démontré d'impact du MMR comparable à celui observé dans des modèles mammifères murins (Usdin et al. 2015, Schmidt et al. 2016). En effet, chez la souris, le MMR est un facteur clef de l'instabilité dans les pathologies à TNRs. Il a notamment été démontré que le complexe MSH2/MSH3 était primordial dans l'instabilité des répétitions GAA/TTC, CAG/CTG et CGG/CTG (Manley et al. 1999, Foiry et al. 2006, Ezzatizadeh et al. 2012, Halabi et al. 2012) Le rôle fondamental de Msh3 et Msh2 sur l'instabilité des TNRs a été notamment démontré au laboratoire dans le modèle murin DMSXL de la DM1. Dans ce modèle, il a été démontré que l'absence de Msh2 induit une disparition quasi complète des expansions en faveur des contractions intergénérationnelles et somatiques (Savouret et al. 2003). Par ailleurs, l'absence de Msh3 diminue les expansions et provoque des contractions intergénérationnelles et est un facteur limitant pour l'instabilité des triplets CTG (Foiry et al. 2006). L'action du MMR semble cependant différente selon la nature des répétitions. En effet, il a été démontré dans un modèle murin de la FRDA invalidé pour les gènes Msh2, Msh3, Msh6 ou Pms2, que le MMR est protecteur contre les expansions (Msh6 et Pms2) et les contractions (Msh2, Msh3) (Ezzatizadeh et al. 2012). Le tableau 9 ci-dessous présente quelques exemples de la démonstration de l'implication du MMR dans l'instabilité des TNRs pathologiques.

Protéine	Pathologie - Modèle	Génotype	Conséquence de l'invalidation de la protéine sur l'instabilité intergénérationnelle	Conséquence de l'invalidation de la protéine sur l'instabilité somatique	Références	
	DM1 - DM300		Contractions paternelles	Diminution des expansions/augmentation	(Savouret <i>et al.</i> 2003,	
				des contractions	Savouret et al. 2004)	
	HD - R6/1		Elimination des expansions	Elimination des expansions	(Manley <i>et al.</i> 1999)	
			Elimination des expansions lors des		(Wheeler <i>et al.</i> 2007	
MSH2	HD - HdhQ11	Msh2 -/-	transmissions paternelles/ Aucun impact sur	Elimination les expansions	Dragileva <i>et al.</i> 2009)	
			transmissions maternelles		- 3	
	FXS - PM		Elimination des expansions	Stabilisation	(Lokanga <i>et al.</i> 2014)	
			Augmente l'instabilité vers les expansions	Diminution des expansions	(Bourn <i>et al.</i> 2012,	
	1 NDA - 1022				Ezzatizadeh <i>et al.</i> 2012)	
	DM1 - DM300	Msh3 -/-	Contractions	Elimination des expansions/	(Foirv <i>et al.</i> 2006)	
		World /		augmentation des contractions	(1 only of an 2000)	
	HD - R6/1	Expression			$(T_{\text{cm}}, c, c, c, l, 2012)$	
MSH3	/BALB/cByJ	réduite de MSH3	ND	Elimination des expansions	(1011e el al. 2013)	
			Aucun impact sur transmissions maternelles/			
	HD - HdhQ11	Msh3 -/-	Biais vers les contractions lors de	Diminution des expansions	(Dragileva <i>et al.</i> 2009)	
			transmissions paternelles			
	FXS - PM	Msh3 - / +	Diminution des expansions	Diminution des expansions	(Zhao <i>et al.</i> 2015)	
	FRDA - YG22	Msh3 -/-	Augmente l'instabilité vers les expansions	ND	(Ezzatizadeh <i>et al.</i> 2012)	

Tableau 7: Exemples de démonstration de l'impact des protéines du MMR dans différents modèles murins (1/2)

Protéine	Pathologie - Modèle	Génotype	Conséquence de l'invalidation de la protéine sur l'instabilité intergénérationnelle	Conséquence de l'invalidation de la protéine sur l'instabilité somatique	Références
MSH6	DM1 - C57Bl/6 × 129/Ola	Msh6 -/-	ND	Augmente l'instabilité	(van den Broek <i>et al.</i> 2002)
	DM1 - DM300	Msh6 -/-	Diminution indirecte des expansions par transmission maternelles	Aucun impact	(Foiry <i>et al.</i> 2006)
	HD - HdhQ11	Msh6 -/+	Augmentation des contractions lors de transmissions paternelle/ Aucun impact sur transmissions maternelles	Aucun impact	(Dragileva <i>et al.</i> 2009)
	FXS - PM	Msh6 - /-	Diminution des expansions	Diminution des expansions	(Zhao <i>et al.</i> 2016)
	FRDA - YG22	Msh6 -/-	Augmente l'instabilité	Diminution des expansions	(Ezzatizadeh <i>et al.</i> 2012)
MLH1	HD - HdhQ11	mlh1 -/-	ND	Stabilisation	(Pinto <i>et al.</i> 2013)
MLH3	HD - HdhQ11	mlh3 -/-	ND	Stabilisation	(Pinto <i>et al.</i> 2013)
PMS2	DM1 - /FVB/N/Dmt-D	Dmo2 /	ND	Diminution des expansions / augmentation des contractions	(Gomes-Pereira <i>et al.</i> 2004)
	FRDA - YG22 -	PIII52 -/-	Augmente l'instabilité	Augmente les expansions, mais diminue l'amplitude de l'instabilité	(Bourn <i>et al.</i> 2012, Ezzatizadeh <i>et al.</i> 2012)

Tableau 8: Exemples de démonstration de l'impact des protéines du MMR dans différents modèles murins (2/2)

Comme le démontrent les données dans des modèles murins du tableau 9, le MMR impacte l'instabilité des triplets à différents stades de ce mécanisme de réparation : au moment de la reconnaissance des lésions par les protéines MSHs, et ensuite *via* les protéines recrutées en aval. Les modèles murins ont été (et sont encore) une source importante d'informations sur les mécanismes de l'instabilité, notamment du MMR. Ces modèles ont ouvert à la voie à des études dans des modèles humains qui ont permis de confirmer les informations prédites par les modèles murins.

III.4.2.3.2. PREUVES DE L'IMPLICATION DU MMR CHEZ LES PATIENTS ET MODELES CELLULAIRES HUMAINS

L'équilibre et la fonctionnalité des complexes MSHs est primordial dans l'instabilité (Keogh et al. 2017). Des modèles de fibroblastes dérivés de patients FRDA, DM1 ont confirmé qu'en absence de MSH2, MSH3, MSH6 l'instabilité somatique était neutralisée (Du et al. 2012, Du et al. 2013, Nakatani et al. 2015). De plus, il a été démontré dans les souris HD et DM1 et en cellules humaines qu'en cas d'invalidation de l'activité ATPasique de Mutsß l'instabilité est biaisée vers les contractions (Tomé et al. 2009, Tome et al. 2013, Keogh et al. 2017), mettant en avant l'impact primordial de ce complexe et plus particulièrement de MSH3 dans l'instabilité des triplets. De plus, lorsque la formation du complexe MSH2/MSH6 est inhibé dans un modèle de cellules humaines HT-1080 avec 800 répétitions CTG, cela favorise l'instabilité. Les auteurs proposent que l'inhibition de la formation de complexes MSH2/MSH6 mène à la formation davantage de complexe MSH2/MSH3, de par la surabondance de MSH2 disponible, ce qui favoriserait l'instabilité (Nakatani et al. 2015). Cependant, il a été observé que, dans des cellules IPSC, que l'invalidation de MSH2/MSH6 stabilise au contraire les triplets GAA (Du et al. 2012). Cette différence peut-être due à la formation de différentes structures secondaires. En effet, les CTG formeront préférentiellement des hairpins qui peuvent être reconnues par MSH2/MSH3 (voir paragraphe III.4.2.3.2.), tandis qu'il est prédit que les répétitions GAA/TTC forment des triplexes. Il est aussi possible que cette différence soit due à l'état de pluripotence des IPSCs, auquel cas il s'agirait d'une démonstration de l'implication de différentes modulations du MMR en fonction de l'état de différenciation des cellules (Du et al. 2013).

Par ailleurs, une étude de cohorte de patients DM1 a confirmé que le polymorphisme de *MSH3* était corrélé à l'instabilité somatique des patients (*Morales et al. 2016*). Récemment, l'étude de polymorphismes de *MSH3* dans des cohortes de patients HD et DM1 a démontré que si un polymorphisme induisait une sous-expression de *MSH3*, l'instabilité somatique des patients est significativement réduite et l'âge d'apparition des symptômes est retardé (Flower *et al.* 2019), confirmant ainsi ce qui avait pu être observé dans les modèles murins auparavant.

Le lien entre l'instabilité des triplets et les protéines du MMR en aval de la reconnaissance des lésions n'a à ce jour pas été démontré dans des modèles cellulaires humains ou chez les patients. Cependant des analyses de GWAS chez des patients HD a démontré un lien entre le polymorphisme de *MLH1*, et la progression de la maladie (Genetic Modifiers of Huntington's Disease 2015, Lee *et al.* 2017), mais le même type d'expérience sur des patients DM1 n'a mis en évidence aucun lien entre des polymorphismes de PMS2 et MLH1 avec l'évolution de la pathologie (Morales *et al.* 2016). Des SNP de *PMS2* ont également été reliés à la progression de la HD et des SCA_{1,2,3,6,7,17} (Bettencourt *et al.* 2016).

III.4.2.3.2. MECANISME DU MMR DANS L'INSTABILITE DES TNRS

Le MMR a été initialement identifié comme un protecteur contre l'instabilité du génome. En effet, dans le cas de cellules tumorales, ce système de réparation est lésé, ce qui provoque l'instabilité des séquences microsatellites (Lynch 1999). Cependant, comme nous l'avons décrit précédemment, il semblerait que, dans les pathologies à TNRs, le MMR normal soit vecteur d'instabilité biaisée vers les expansions. Les mécanismes impliqués dans l'instabilité des microsatellites des cancers et des maladies à TNRs seraient donc différents. Dans ce chapitre, nous exposerons comment un MMR normal peut favoriser l'instabilité des triplets répétés.

-MMR et structures secondaires

Le recrutement du MMR sur les TNRs pathologiques est lié à la formation de structures secondaires. En effet, dans un modèle *in vitro* à base d'oligonucléotides et de protéines de la machinerie MMR humaine purifiées, le complexe MSH2/MSH3 fixe les mésappariements A-A et T-T de la tige des hairpins d'oligonucléotides CAG ou CTG. Un modèle proposé est que le MMR y reste captif, et son l'activité ATPasique est inhibé, ce qui empêcherait l'intervention réparatrice du MMR (Owen *et al.* 2005, Lang *et al.* 2011). Les hairpins seraient alors stabilisées, car rendues inaccessibles aux mécanismes de réparation par le complexe MSH2/MSH3 invalidé. Les hairpin sont alors intégrées à l'ADN au cours de la réplication ou de la réparation, ce qui provoque des expansions (voir figure 19).



Figure 19: Illustration de la stabilisation d'une hairpin par le complexe MSH2/MSH3.

Par ailleurs, les boucles extra-hélicoïdales de 20 répétitions CAG ou CTG, ou de plus petites boucles de 10 répétitions CTG, CAG, CCG ou CGG, ne sont efficacement réparées que si le MMR les prend en charge en 5' : les hairpins sont éliminées, ce qui aboutit à une contraction des répétitions. En revanche, lors de reconnaissances en 3' par MSH2/MSH3, les structures en hairpins sont résistantes à la réparation. Il est possible que cette différence de traitement soit due à une séquestration du complexe MMR dans des structures secondaires qui se forment en cas de prise en charge en 3' (voir (lyer *et al.* 2015) pour revue). Une autre étude dans des tissus de modèles murins HD a quant à elle plutôt suggéré que lors de la transcription, la structure des répétitions de CAG soit dynamique avec l'ouverture de la chromatine qui favorise l'accès les mécanismes de réparation: le MMR et BER agissent alors de concert sur ces structures (Goula *et al.* 2009).

-MMR et BER

Le MMR et le BER interagissent en cas de répétitions de triplets, notamment en cas de lésion oxydative de répétitions GAA ou CAG. Le complexe MSH2/MSH3 se fixe à la séquence répétée et force le passage de la polymérase ß du BER qui échoue à répliquer les TNRs correctement. La polymérase ß va alors « sauter » des séquences ce qui provoquera des délétions et laissera perdurer des structures normalement intermédiaires. Ces délétions et structures anormalement stables seront ensuite reconnues par le complexe MSH2/MSH3, qui va soit stimuler la voie du MMR, soit forcer la polymérase ß à répliquer les TNRs mésappariés, par la synthèse de précurseurs flaps, vecteurs d'expansion (Goula *et al.* 2012, Lai *et al.* 2016). Par ailleurs, il a été démontré que lorsque MSH2/MSH3 fixe une hairpin, le MMR recrute l'ADN polymérase ß du BER qui se sert de l'hairpin comme d'un « primer » d'extension, ce qui provoquera également une expansion de triplets (Guo *et al.* 2016).

À travers ce chapitre, nous avons illustré l'importance du mécanisme de réparation dans l'instabilité des triplets répétés. Ce mécanisme est d'autant plus fondamental dans l'instabilité des TNRs qu'il intervient dans les cas de cassures, mais aussi dans la réplication et, comme nous le verrons dans le paragraphe suivant, le MMR intervient aussi au cours de la transcription, et plus particulièrement dans le cas de la transcription des triplets répétés.

III.4.3. LA TRANSCRIPTION

La transcription eucaryote dépend de 3 ARN polymérases, et seule l'ARN polymérase II (RNAPII) synthétise les ARN messagers (ARNm). La transcription débute après la reconnaissance d'une zone promotrice centrale où se fixera la RNAPII (la TATA box), généralement localisée à une trentaine de nucléotides de la région codante. L'efficacité de la transcription dépendra des régions modulatrices (« enhancer » ou « silencer ») dans le voisinage du site de fixation de la RNAPII et des facteurs de transcription qui rejoindront la polymérase et qui l'aideront dans sa progression : facteurs de transcription, résolveurs de structures de la chromatine, modificateurs d'histones etc... (voir figure 20)



Figure 20: Aperçu général de la Transcription eucaryote.

(a) Représentation schématique de la région initiatrice de la transcription eucaryote et de l'initiation de la transcription. Le Médiateur permet de transférer le pouvoir activateur des enhancers (ou répresseur des silencers) sur le complexe transcriptionnel, les enzymes modificatrices d'histones et de remodelage permettront une meilleure ouverture de la double hélice. (b) Schématisation de la progression de l'élongation transcriptionnelle. L'ARN polymérase progresse le long de l'ADN matrice en synthétisant l'ARNm complémentaire par jointure de ribonucléosides triphosphates libres. Un court segment ADN/ARN est formé dans la polymérase, ce qui pourra former des R-loops si la structure persiste après le passage de la polymérase. (c) Repliements possibles de la chromatine suite à la fixation d'une protéine sur la double hélice. L'ADN prendra une structure superenroulée positive ou négative qui viendra influencer le sens de progression de la protéine. (Figure adaptée et traduite d'après Bruce Alberts Alexander Johnson Julian Lewis Martin Raff Keith Roberts Peter. Molecular Biology of the Cell. 5th ed.; p370-380. Edited by Garland Science. 2008.)

L'ARN polymérase se déplace le long de la double hélice ouverte, sur le brin d'ADN matrice et synthétise son ARNm complémentaire en liant des ribonucléosides triphospates entre-eux. Durant le processus, l'ADN est donc transitoirement simple brin. Au cours du

déplacement de la polymérase, l'ADN se déformera, formant des structures super-enroulées à cause de la tension exercée par la transcription (voir figure 20). De plus, la chromatine subira un enroulement/désenroulement dynamique au niveau des nucléosomes, ce qui favorise la formation de structures secondaires. L'ARNm synthétisé court dans le sillon du complexe transcriptionnel, et en ressort par l'arrière au fur et à mesure que le complexe progresse (voir figure 20). L'hélice d'ADN étant ouverte, l'ARNm et son brin d'ADN complémentaires peuvent parfois former des complexes hybrides appelés R-Loop (voir (Aguilera *et al.* 2012) pour revue). Lorsque le complexe transcriptionnel reconnait une séquence terminatrice, celui-ci se disloque, libérant (1) l'ADN qui reformera normalement une séquence double brin en double hélice bêta ; (2) l'ARNm qui sera ensuite pris en charge pour des modifications post-transcriptionnelles (pose de la coiffe, de la queue polyA...) avant d'être exporté du noyau.

A travers cette brève description de la transcription eucaryote, il est déjà des mécanismes identifiables comme vecteurs d'instabilité du génome. À travers les paragraphes suivants, nous décrirons comment la transcription peut-être source d'instabilité dans le génome, et plus précisément dans les TNRs qui sont déstabilisés par la transcription (voir (Lin *et al.* 2009) pour revue).

III.4.3.2. NIVEAU DE TRANSCRIPTION

Les gènes sont différemment transcrits en fonction du type cellulaire. Ce nivellement de la transcription est notamment médié par le repliement de la chromatine et les marques épigénétiques. Il a été démontré dans des modèles de cellules CHO, que le niveau de transcription était corrélé avec le degré de recombinaison homologue intra-chromosomique (Gottipati *et al.* 2008). Des modèles de cellules humaines tumorales ont par ailleurs prouvé que l'augmentation du niveau de transcription menait à l'accumulation d'hybrides ADN/ARN, dérégulant notamment les fourches de réplication et induisant des cassures de l'ADN (Kotsantis *et al.* 2016).

Par ailleurs, il a été démontré que les répétitions de CGG associées à une pré-mutation de l'X Fragile ainsi que les répétitions CAG de la HD et CTG de la DM1 doivent être transcrites pour former des expansions ((Lin *et al.* 2009, Adihe Lokanga *et al.* 2014). En comparaison, il est intéressant de constater que, chez l'Homme, la mutation complète menant au FXS n'est pas transcrite et que la répétition de CGG demeure stable au cours du temps. Lorsque la mutation est transcrite *in vitro*, les auteurs observent des expansions (Wöhrle *et al.* 2001). Il a été démontré que la transcription était un acteur important (voire indispensable) pour l'instabilité des triplets répétés, cependant, à ce jour, aucune corrélation entre le niveau de transcription et l'instabilité des triplets chez les eucaryotes (voir (Dion *et al.* 2009, Lin *et al.* 2009, Lokanga *et al.* 2013, Usdin *et al.* 2015) pour revues).

III.4.3.2. PAUSES DE LA TRANSCRIPTION

L'élongation au cours de la transcription est régulièrement interrompue, sur tout le segment transcrit aussi bien lors de la transcription sens qu'antisens (voir (Imashimizu *et al.* 2018). Lorsque des hybrides ADN/ARN sont détectés, cela retarde la translocation de l'ARN polymérase et l'intégration des derniers ribonucléosides dans l'ARN néosynthétisé (Bochkareva *et al.* 2012). Certaines séquences nucléotidiques (riches en G sur le brin non-matrice) induisent la pause de l'élongation de façon ubiquitaire, les cassures et la compétition entre les protéines de la réparation et de la transcription sur l'ADN sont aussi incompatibles avec l'élongation. Cela provoque notamment l'isomérisation du complexe d'élongation en une structure « en pause » qui est inactive mais l'ARN polymérase reste en place (Landick 2006) : une pause au niveau des zones promotrices permettent notamment de garder la double hélice ouverte et de garantir un haut niveau de transcription sur le gène concerné.

Cependant, une pause anormalement longue de l'ARN polymérase va mener à l'instabilité du génome par compétition avec les fourches de réplication de l'ADN. Normalement, ces deux mécanismes ne peuvent pas cohabiter chez les eucaryotes grâce aux séquences isolatrices induisant une pause de la fourche de réplication, mais aussi par une séparation temporelle entre ces deux métabolismes (voir (Helmrich *et al.* 2013, Hamperl *et al.* 2016) pour revue). Cependant, quand il arrive malgré tout ces deux métabolismes cohabitent, cela abouti à une collision entre la machinerie de réplication et de transcription, ce qui forme des R-loops, des boucles d'ADN hors hélices... qui seront reconnues par les mécanismes de réparation (Helmrich *et al.* 2013).

Des études *in vitro* par utilisation de plasmides ont démontré que des expansions de 17, 50, et 255 triplets CTG ne stoppent pas l'avancée de la RNAPII (ARN Polymérase de type II) II, mais ralentissent significativement sa progression dans les répétitions. En effet, la RNAPII est alors sous une forme « incompétente », et un pic de ralentissement est observé aux premier, deuxième, sixième et neuvième triplets, indépendamment de la taille des répétitions (Parsons *et al.* 1998). Un autre modèle de transcription *in vitro* avec des protéines d'extraits nucléaires de cellules HeLa a démontré que la RNAPII était en pause en présence de 20 répétitions CTG ou CAG à cause des hairpins présentes sur les répétitions, que ce soit sur le brin transcrit ou non transcrit (Salinas-Rios *et al.* 2011). Les expansions GAA/TTC inhibent la transcription *in vitro* de par le superenroulement de l'ADN et les structures triplexes que prennent ces répétitions (Grabczyk *et al.* 2000).

III.4.3.3. LA TRANSCRIPTION BIDIRECTIONNELLE

Il est estimé que 30% des gènes humains sont transcrits dans l'orientation sens et antisens par la RNAPII. La plupart du temps, le niveau de transcription « sens » est plus élevé que celui de la transcription antisens. De plus, chez l'Homme, le transcriptome antisens varie en fonction du type cellulaire (He *et al.* 2008). Si 2 promoteurs sont suffisamment proches l'un de l'autre et s'ils induisent des transcriptions convergentes, cela mène à la collision frontale des machineries de transcription qui s'arrêtent sur la zone de collision et collapsent. La collision entraînerait des cassures, la formation de structures secondaires, des sur-enroulement de l'ADN et favorise l'hybridation ADN/ARN (R-Loops) qui peuvent être reconnues par les mécanismes de réparation (voir (Lin *et al.* 2011, Kim *et al.* 2012) pour revue).

La transcription bidirectionnelle est présente sur de nombreux *loci* humains de maladies à TNRs : DM1, FXS, SCA7,8, HD, HDL2, FRDA par exemple (voir (Usdin *et al.* 2015) pour revue). Dans le cas du gène *DMPK* de manière générale, il a été démontré que la transcription bidirectionnelle mène à une production d'ARN antisens moins abondante que celle d'ARN sens. Ces transcrits *DMPK* antisens correspondent à de nombreux petits intermédiaires de transcription, contenant ou pas les répétitions. Les ARNm antisens sont convertis en siARN (ARN silencer), qui, en fixant leur ADN complémentaire iraient stimuler son hétérochomatisation par marque épigénétique (méthylation), empêchant ainsi la transcription (Cho *et al.* 2005). Ce mécanisme est commun aux allèles *DMPK* mutant et sain. Les patients présentent davantage de transcrits antisens que dans la population générale, mais il demeure moins de transcription dans l'orientation antisens que sens. De plus, les auteurs ont également noté que la quantité de transcrits antisens dans les noyaux de cellules de patients DM1 augmentait en même temps que l'aggravation des symptômes (Gudde *et al.* 2017).

Il a été proposé que l'augmentation anormale de la quantité de transcrit antisens dérégulerait le système de siARN en induisant une sur-méthylation aberrante ainsi que des dommages par stress oxydatifs (Budworth *et al.* 2013). Par ailleurs, la présence des répétitions anormales dans l'ADN des patients DM1 est propice à la formation de structures secondaires qui vont d'autant plus fragiliser la séquence répétée par hétérochromatisation. Cette condensation de l'ADN est propice à la formation de structures secondaires au niveau des répétitions CTG et des séquences environnantes, ce qui favorise l'instabilité des TNRs. Les mécanismes exacts par lequel la transcription bidirectionnelle pourrait influencer l'instabilité des TNRs ne sont à ce jour pas clairement caractérisés. Cependant, plusieurs modèles cellulaires confirment que l'induction de la transcription bidirectionnelle favorise l'instabilité des

75

TNRs, démontrant l'importance du phénomène dans l'instabilité des triplets répétés ((Nakamori *et al.* 2010), voir pour revues (Lin *et al.* 2012, Usdin *et al.* 2015)).

III.4.3.4. LES R-LOOPS

Les R-loops sont des structures non bêta à 3 brins, composée d'une séquence hybride ARN/ARN et d'un brin d'ADN simple brin (voir figure 21). D'abord considérées comme des intermédiaires de la transcription, de la réplication, et/ ou des acteurs de l'immunologie prokaryotes via Crispr (Ivančić-Baće *et al.* 2012), les R-loops sont aujourd'hui reconnues comme de véritables facteurs modulateurs de l'expression des gènes ainsi que de l'instabilité du génome.



Figure 21: Schématisation d'une R-loop D'après Ivančić-baće I et al. 2012

<u>APERÇU DE LA REPARTITION DES R-LOOPS DANS LE GENOME HUMAIN</u>

Les R-loops sont des structures conservées au sein des espèces à ADN double brin. Elles peuvent être notamment le résultat de l'activité des ARN polymérases I, II et III. Des expériences d'immunoprécipitation couplées au séquençage haut débit (DRIP-seq) ont permis de cartographier les R-loops dans le génome humain. Ainsi, il a été démontré que les R-loops sont présentes sur 3-5% du génome, avec une sur-représentation au niveau des régions riches en GC, notamment dans les promoteurs et les régions terminales: 59% des gènes humains présentent des R-loop (Aguilera *et al.* 2012, Jenjaroenpun *et al.* 2016), un aperçu plus précis de la répartition des R-loops en fonction de la classe des gènes est proposée dans le tableau 10. Les promoteurs et terminateurs formant des R-loops sont particulièrement touchés par les phénomènes de pauses de la transcription et sont enrichis en marques épigénétiques d'ouverture de la chromatine. Cette répartition des R-loops dans le génome en corrélation avec les marques épigénétiques est en adéquation avec le rôle des R-loops dans la régulation positive de l'expression (Chédin 2016, Sanz *et al.* 2019).

Gènes codant pour	Pseudogènes	Gène codant de long	Gènes codant de courts
une protéine		ARN non traduits	ARN non traduits
75%	28%	60%	23%

Tableau 9: Ratio des gènes humains positifs pour la formation d'R-Loops selon leur classe.D'après (Aguilera et al. 2012, Jenjaroenpun et al. 2016)

• DYNAMIQUE DES R-LOOPS

Les R-loops ont initialement été observées sur une longueur de quelques nucléotides au cours de la transcription. Ces structures se dissocient normalement rapidement après le passage de la RNAPII. La formation d'R-loops plus stables est liée à des séquences dites « propices » (riches en GC) et à différents métabolismes de l'ADN. À noter cependant que la formation d'R-loops sur les zones propices n'est pas automatique (voir (Aguilera *et al.* 2012, Belotserkovskii *et al.* 2018, Crossley *et al.* 2019) pour revue). La survenue d'R-loops est beaucoup plus probable si le gène concerné subit plusieurs passages de la machinerie de transcription, et si le site d'initiation de la transcription est proche des séquences propices à la formation de R-loops. Les R-loops peuvent également se former à la suite d'une collision entre les machineries de réplication et de transcription. Les fourches de réplication et machineries de transcription peuvent entrer en collision frontale ou alors l'une des machineries peut en rattraper une autre qui serait « en pause ». Lors d'une collision frontale, la formation de R-loops augmente. Il est supposé que cela soit dû à l'arrêt de l'ARN polymérase qui induirait un mauvais traitement des fragments d'Okazaki : l'amorce ARN n'est pas dégradée et reste hybridée au brin répliqué indirect.

La structure des R-loops est un intermédiaire entre la forme bêta (configuration canonique de l'ADN double brin) et la forme alpha (forme canonique des ARN doubles brin). La liaison ADN/ARN est beaucoup plus forte que la liaison ADN/ARN. Cependant, cette stabilité dépendra de plusieurs paramètres :

- La topologie de la séquence : par exemples, les super-enroulement négatifs, les structures en hairpin, G-quadruplexes et triplexes sur le brin non transcrit, stabilisent les R-loops.
- La nature de la séquence : les liaisons permettant les R-loops les plus stables sont les liaisons rG/dC, et les moins stables seront les liaisons rU/dA.
- La séquestration du brin non transcrit par un ligand qui par encombrement empêche la libération de l'ARN.

Les R-loops sont des éléments parfois indispensables à la régulation de l'expression de gènes, parfois elles sont accidentelles. Cependant, en toutes circonstances, toutes les Rloops doivent finir par être détruites : il a été démontré qu'en absence des hélicases spécifiques des liaisons ADN/ARN et des RNses H, que le génome présente une hausse de l'instabilité et des cassures du génome (Skourti-Stathaki et al. 2014, Santos-Pereira et al. 2015, Sollier et al. 2015). La présence d'R-loops est un phénomène très régulé par différentes voies. Par exemple, il existe des familles d'hélicases (RNA-DNA helicases, SETX, Aquarius, DHX9, Topoisomérase 1...) capables de déstabiliser des R-loops au cours de la réplication et/ou de la transcription. L'action des hélicases rendra les hybrides ADN/ARN accessibles aux RNases H1 et H2 qui dégraderont l'hybride ADN/ARN. Les mécanismes de réparation de l'ADN comme le TCR, ont des nucléases XPG et XPF qui clivent les R-loops causées par la transcription. Une R-loop trop persistante mènera à une dérégulation du métabolisme de l'ADN (collision, recrutement de machinerie de réparation aberrant...). Plus une R-loop est stable, plus elle sera résistante à sa destruction et aura tendance à être connue comme une erreur par les mécanismes de réparation (McIvor et al. 2010, Helmrich et al. 2013, Belotserkovskii et al. 2018, Freudenreich 2018, Crossley et al. 2019).

<u>R-LOOPS ET REGULATION DE L'EXPRESSION DES GENES</u>

Les R-loops peuvent stimuler ou réprimer l'expression de gènes. En effet, les R-loops rendent par exemple les promoteurs de la transcription inaccessibles aux mécanismes de méthylation notamment au niveau des séquences CpG (ce qui favorise donc l'expression), mais les R-loops peuvent aussi ralentir voire bloquer la progression de la machinerie de transcription (Groh *et al.* 2014). Au niveau des régions terminatrices, les R-loops permettent une pause de la RNAPII à proximité des sites de terminaison. Cet arrêt permet de rendre accessible l'ARN néosynthétisé aux facteurs de polyadénylation et de terminaison de la transcription. De plus, ces pauses évitent les collisions des transcriptions simultanées de gènes proches à transcriptions convergentes. Enfin, les R-loops combinées à une transcription antisens peuvent former des complexes ARN double sens qui vont recruter des marques épigénétiques d'hétérochomatisation, réduisant ainsi l'expression. (voir (Ginno *et al.* 2012) (Belotserkovskii *et al.* 2018, Crossley *et al.* 2019) pour revues).

EXEMPLES D'IMPLICATION DES R-LOOPS DANS L'INSTABILITE DU GENOME

L'implication des R-loops dans l'instabilité du génome la plus référencée est l'hyperrecombinaison nécessaire à la commutation de classe des immunoglobuline (Ig). Chez les mammifères, la zone de commutation s'étend sur plusieurs kilobases, avec une succession d'unités de séquences transcrites mais non traduites, avec leurs propres promoteurs. Cette zone est répétitive et riche en guanines : elle est donc propice à la formation d'R-loops et la présence de ces dernières sur cette zone a été démontré chez l'Homme. La commutation de classe nécessite une activation induite par désaminase (AID) qui est recrutée sur les cytosines de l'ADN simple brin impliqué dans la R-loop (Roy *et al.* 2008) : ce mécanisme est indispensable à la commutation de classe des lg.

Les R-loops sont donc vectrices d'instabilité physiologique dans le cas des immunoglobulines. Mais les R-loops sont aussi impliquées dans les lésions « accidentelles » de l'ADN. En effet, la surexpression de la RNAse H (éliminant les R-loop) dans des cellules humaines diminue significativement les cassures double brin; à l'inverse, en cas de déplétion de la RNase H ou des hélicases à ARN, le génome accumule les lésions notamment doubles brins. En formant des structures secondaires non canoniques et en induisant des pauses de la réplication et de la transcription, les R-loops provoquent l'instabilité du génome par les mêmes mécanismes que ceux des structures secondaires décrits dans les chapitres précédents : ADN simple brin fragilisé, réparation anormale, dérapage des polymérases, cassures etc...(voir (Aguilera *et al.* 2012, Belotserkovskii *et al.* 2018) pour revues).

Par ailleurs, le rôle primordial des R-loops a récemment été mis en avant dans la réparation par recombinaison. Dans un modèle de cellules humaines, il a été démontré que les R-loops s'accumulent au voisinage de cassures doubles brins, sous l'impulsion de DROSHA (ribonucléase de type III). DROSHA permettrait le clivage des R-loop en miARN (micro ARN) qui faciliteraient la recombinaison homologue et non homologue en servant de modèle en cas de ressection des brins d'ADN. Si le mécanisme en soit n'est à ce jour pas totalement élucidé, il est prouvé qu'en cas d'inhibition de DROSHA, les R-loops ne se forment plus autour des cassures double brin, et les réparations par recombinaison sont moins efficaces (Lu *et al.* 2018).

De nombreuses pathologies humaines sont influencées par l'instabilité du génome provoquée par une mauvaise régulation des R-loops (Crossley *et al.* 2019). Par exemple, certains cancers du sein mènent à la séquestration de BRCA1 qui élimine normalement les R-loops dans les régions terminatrices. Cela provoque une instabilité du génome participant à la carcinogénèse ; autre exemple avec l'anémie de Fanconi dont les facteurs du complexe FANCD2 sont déficients, ce qui stimule la formation d'R-loops et les interventions de TCR ; dans le syndrome d'Aicardi-Goutière, (pouvant être causé par différentes mutations types insertion/délétion sur n'importe quel gènes codant pour les complexes TREX1 (AGS1), la RNAse H2 endonucléase, la triphosphohydrolase SAMHD1 et le gène ADAR1) le niveau global d'R-loop est augmenté, en corrélation avec une diminution du niveau de méthylation. Nous verrons dans le paragraphe suivant que l'instabilité des triplets répétés dans les maladies humaines à mutations dynamiques est aussi particulièrement impactée par les R-loops.

LES R-LOOPS DANS LES MALADIES A TRIPLETS REPETES

Les *loci* des maladies à TNRs sont particulièrement propices à la formation d'R-loops. En effet, les triplets répétés CNG sont intrinsèquement des séquences « propices » à la formation de R-loops (car riches en GC) et leurs séquences flanquantes, en amont et en aval, ont un pourcentage en GC significativement plus élevé que dans le reste du génome ((McIvor *et al.* 2010) et figure 22).



Figure 22: Schématisation de la formation co-transcriptionnelle d'R-loops dans les répétitions CNG.

Le pourcentage en GC des séquences flanquantes de la majorité des loci à TNRs instables est significativement plus élevé que dans le reste du génome (moyenne représentée en pointillés). La plupart des TNRs instables sont propices à la formation de structures secondaires (ici exemple des hairpins en rouge) qui facilitent l'hybridation entre l'ARN (en rouge et bleu) à son brin transcrit complémentaire. La majorité des loci liés à des pathologies à triplets ont une transcription bidirectionnelle : les transcrits jaunes et bleus sont issus de différents passages de transcription en sens et antisens. Adaptation et traduction d'après McIvor et al. 2010.

Des expériences de transcription *in vitro* et en cellules humaines ont démontré que les R-loops pouvaient se former sur les répétitions CAG, CTG, GAA, CGG et CCG (voir (McIvor *et al.* 2010, Reddy *et al.* 2011, Richard *et al.* 2017, Crossley *et al.* 2019) pour revue et tableau 11). La formation d'R-loops est dépendante de la nature des triplets. En effet, la transcription *in vitro* de répétitions TTC ne forme pas d'R-loop, contrairement à son brin complémentaire GAA (Reddy *et al.* 2011). Ou encore, dans des modèles cellulaires FRAXA et DM1, il a été démontré que les transcrits rCGG et rCAG formaient plus fréquemment des R-loops que leurs transcrits complémentaires rCCG et rCUG (Lin *et al.* 2010, Reddy *et al.* 2011).

Maladie	Gène/ Répétition	Conséquences des R-loops	Références
DM1	DMPK / CTG	Stimule l'instabilité des triplets	(Lin <i>et al.</i> 2010)
HD	HTT / CAG	Stimule l'instabilité des triplets	(Reddy <i>et al.</i> 2011)
FXS	FMR1 / CGG	Empêche le passage de RNAPII et invalide l'expression du gène par hétérochromatisation	(Loomis <i>et al.</i> 2014)
FRDA	FXN / GAA	Empêche le passage de RNAPII et invalide l'expression du gène par hétérochromatisation	(Groh <i>et al.</i> 2014)

Tableau 10: Exemples d'implication des R-loops dans les maladies à TNRs en cellules humaines.

Des expériences en cellules humaines ont démontré que d'invalidation de la RNase H stimulait l'instabilité des triplets CTG/CAG, démontrant ainsi que les R-loops détectées dans les triplets étaient vectrices d'instabilité (Lin *et al.* 2010). Les auteurs ont également prouvé que le sens de la transcription formant des R-loops à travers les triplets CAG/CTG influence les biais de l'instabilité : la formation d'R-loops par transcription bidirectionnelle stimule d'avantage l'instabilité que lors de la formation d'R-loop par transcription d'un seul des 2 brins sens ou antisens. De plus, l'induction d'R-loop rCUG/dCAG (le brin porteur de CTG est dans cette configuration simple brin) provoque un biais vers les contractions, tandis que l'induction d'R-loop rCAG/dCTG biaise l'instabilité vers les expansions (Reddy *et al.* 2011).

Les R-loops dans les TNRs influencent l'instabilité des triplets en collaboration avec les mécanismes de réparation (voir (Freudenreich 2018) pour revue). Dans un modèle de levure, il a été démontré que les répétitions CAG/CTG étaient sujettes à la désamination des cytosines sur le simple brin d'ADN de l'R-loops qui va être reconnue pas le BER, ce qui induit l'instabilité des triplets ((Su *et al.* 2017) et voir chapitre III-4.3.2). Dans le même modèle, les auteurs ont démontré que le MMR provoquait des contractions liées aux R-loops. Un autre modèle de levure de la FRDA a mis en évidence que les R-loops présentes sur les répétitions GAA favorisent l'instabilité par induction de cassure liée à la réplication et au recrutement du BIR (Neil *et al.* 2018).

La découverte du métabolisme des R-loops a rajouté un nouvel élément de régulation de l'expression et de la stabilité du génome, notamment des TNRs. Ces structures sont un élément supplémentaire dans le puzzle de l'instabilité et font déjà l'objet de recherche à vocation thérapeutique dans le FXS et FRDA (Groh *et al.* 2014): nous discuterons des approches thérapeutiques dédiées aux maladies à TNRs dans le chapitre suivant.

III.5. CONCLUSION DE CHAPITRE

L'instabilité des triplets répétés est un pilier central des maladies à TNRs. À travers ce chapitre, nous avons pu rendre compte de la multitude de mécanismes qui influencent l'instabilité des TNRs : facteurs propres à la séquence d'ADN (structures secondaires, natures des répétitions et des séquences environnantes, modulation par le métabolisme (Réplication, Réparation, Transcription...). De nombreux modèles procaryotes et eucaryotes ont permis de mieux comprendre les mécanismes de l'instabilité des TNRs, cependant certaines parts d'ombres subsistent. Par exemple, pourquoi les *loci* à TNRs mutants ne sont-ils pas traités de la même façon que les autres zones à répétitions par le MMR?

Cependant, l'avancée dans la compréhension des mécanismes des maladies à triplets répétés *via* différents modèles a permis d'élaborer des stratégies thérapeutiques, et permettront certainement d'en développer davantage à la lumière des nouvelles données rendues accessibles par ces outils de laboratoire.

IV. STRATEGIES THERAPEUTIQUES POUR LA DM1

Le développement de modèles murins a particulièrement contribué à la compréhension de la physiopathologie et l'instabilité des TNRs. Ces modèles ont également largement contribué au développement des thérapies ainsi qu'aux tests précliniques. Ces modèles ont fait l'objet de nombreuses revues (quelques exemples dans les références suivantes (Ingram *et al.* 2012, Dandelot *et al.* 2017, Braz *et al.* 2018) Dandelot *et al.* est disponible en Annexe 2 de ce manuscrit). Dans ce chapitre, nous nous intéresseront à des exemples illustrant la variété de stratégies thérapeutiques pouvant traiter différents aspects de la DM1. Nous clôturerons ce chapitre par un bref aperçu sur les essais cliniques.

IV.1. APPROCHES MEDICAMENTEUSES

Les thérapies par replacement de médicaments présentent l'avantage d'avoir déjà été caractérisés pour l'Homme (dose efficace, toxicité, effets secondaires...) et de bénéficier d'une autorisation de mise sur le marché. Quelques molécules ont déjà prouvé leur efficacité dans différents modèles de la DM1. Certaines d'entre-elles diminuent la transcription des répétitions mutantes, d'autres compensent la séquestration de MBNL1, détruisent les *foci* ou corrigent directement des voies de régulations atteintes dans la DM1 (voir (Lopez-Morato *et al.* 2018) pour revue).

IV.1.1. DIMINUTION DE L'EXPRESSION DE DMPK ET CONSEQUENCES

Il a été démontré dans des modèles murins que l'invalidation de *DMPK* n'avait qu'une faible incidence sur le phénotype normal (Jansen *et al.* 1996, Berul *et al.* 1999, Carrell *et al.* 2016). C'est pourquoi certaines pistes thérapeutiques se sont orientées vers une **diminution de l'expression de** *DMPK* pour limiter la production d'ARN toxiques. La **pentamidine** et les antibiotiques associés sont normalement utilisés dans les cas de pneumonies. Il a été démontré que ces molécules pouvaient réduire l'expression de *DMPK* dans différents modèles de cellules humaines et murins. De plus, la pentamidine fixe et réduit la présence de *foci,* libère MBNL1, et restaure l'épissage des pré-ARNm du récepteur à l'insuline et de la troponine T (Warf *et al.* 2009). **L'actinomycine D** (ActD) est un antibiotique utilisé en tant qu'anticancéreux. Cet intercalant de la double hélice d'ADN provoque des cassures, inhibe la transcription et stoppe la prolifération cellulaire. Dans le cas de la DM1, l'ActD réduit la transcription de *DMPK*, diminue la présence de *foci,* libère MBNL1 et rétabli *in vivo* l'épissage complet du canal chlore murin, et partiellement celui d'autres ARNm (Siboni *et al.* 2015).

IV.1.2. MBNL1 HORS DES FOCI

La séquestration de MBNL1 dans les foci est un phénomène crucial dans l'aspect « spicéopathique » de la DM1. Certaines molécules comme le Phénylbutazone (PBZ, un antiinflammatoire) induisent la surexpression de MBNL1 dans un modèle murin de la DM1 HSA (voir (Gomes-Pereira et al. 2011) pour revue sur les modèles murins DM1). L'injection de PBZ résulte en (1) la restauration partielle de l'épissage de certaines cibles de Mbnl1, (2) la diminution de la colocalisation de MBNL1 avec les foci d'ARNm (Overby et al. 2018). Une autre possibilité pour relarguer MBNL1 dans le nucléoplasme est d'intégrer un compétiteur à sa fixation aux ARNm mutants. L'érythromycine (un antibiotique fixant l'ARN) par exemple, délocalise MBNL1 des foci, et rétabli aussi l'épissage de certains ARNm atteints dans la DM1 (Nakamori et al. 2016) Cependant, la délocalisation ou la surexpression de MBNL1 peut à terme devenir délétère. En effet, MBNL1, en plus de son activité de régulateur de l'épissage alternatif, sert de senseur de structures secondaires aberrantes des ARNm, et permet leur correction. Si MBNL1 ne fixe plus aussi efficacement ces structures (hairpins sur CUG allongés, ou accidentelles de manière générale), les ARNm mal repliés vont s'accumuler dans les noyaux, sans être traités par les mécanismes protecteurs de la cellule, et finiront par être toxiques (voir (Thornton et al. 2017, Lopez-Morato et al. 2018) pour revues).

IV.1.3. REGULATION DE LA PHOSPHORYLATION - IMPACT SUR CELF1

Il est connu que différentes kinases sont impactées dans la DM1 (GSK3ß, CDKs, AKT, PKC...). Cette dérégulation mène notamment à **l'hyper-phosphorylation de CELF1**, reconnue comme en partie responsable de défauts d'épissages dans la DM1. Des analyses de screening et d'hybridation *in situ* ont permis d'identifier des modulateurs de phosphatases et de kinases pouvant potentiellement servir à des fins thérapeutiques. Parmi elles, Ro 31-8220 et le lithium.

Ro 31-8220 élimine les *foci*, réduit la fixation de MBNL sur les *foci* et limite l'hyperphosphorylation de CELF1 en inhibant la PKC. L'usage de Ro 31-8220 améliore le phénotype chez la souris HSA, mais le mécanisme exact du mode d'action de Ro 31-8220 reste à élucider. **Le lithium** quant à lui se montre efficace dans l'inhibition de GSK3ß. Or, GSK3ß est une kinase anormalement stable dans la DM1, et est en partie responsable de l'hyperphosphorylation aberrante de CELF1. En inhibant GSK3ß, le lithium restaure la bonne phosphorylation de CELF1. D'un point de vue phénotypique, les souris DM1 traitées au lithium dans les muscles squelettiques recouvrent leur force musculaire (voir (Thornton *et al.* 2017, Lopez-Morato *et al.* 2018) pour revues).

IV.2. THERAPIE PAR L'ARN

Les foci d'ARNm *DMPK* mutants sont la cause de la perte de fonction de MBNL1 et de la dérégulation de CELF1. De nombreuses approches thérapeutiques pour tenter de contrer la formation de *foci* et ainsi restaurer le bon métabolisme des ARN ont été envisagées (voir figure 23 et (Overby *et al.* 2018) pour revue).



Figure 23: Approches thérapeutiques visant à éliminer les foci d'ARNm mutants DMPK.

(a) Situation pathologique dans la DM1. Les ARNm DMPK mutant se replient sous forme d'hairpin et s'agrègent dans le noyau en foci. Cette situation dérégule les protéines MBNL1 et CELF1, ce qui provoque des troubles du métabolisme d'autres ARN. (b) Stratégies thérapeutiques pour détruire les ARNm mutants. Ces stratégies comprennent des ARN interférents, des (d)Cas9 ciblant l'ARN, et des Oligonucléotides Antisens (ASOs) D'après Overby et al. 2018.

Les ARN interférents (ARNi) sont des séquences d'ARN simples ou double brin qui entrent en interférence avec un ARNm cible. Le complexe ARN double brin sera ensuite dégradé *via* le complexe RISC. Sobczak et *al.* ont démontré dans un modèle murin DM1 HSA l'efficacité d'ARNi délivrés par électroporation ou injection intramusculaire et ciblant spécifiquement les expansions CUG. En effet, les souris traitées présentaient 70 à 80% d'ARNm mutants en moins, et le transcrit normal n'était pas affecté. Les auteurs ont observé une diminution significative de la présence de *foci*, et l'épissage de certains ARNm était restauré notamment pour *CLCN1* (Sobczak *et al.* 2013). Une autre étude basée sur des ARNi

spécifiquement dirigés vers les hairpins CUG et apportés par vecteur rAAV *via* des injections intra-veineuses, a également présenté une efficacité dans la réduction du nombre de *foci* et dans la restauration de l'épissage (Bisset *et al.* 2015).

Les oligo-nucléotides antisens (ASO) sont ici une séquence ADN complémentaire aux ARNm *DMPK* mutant. L'hybridation ADN/ARN neutralise l'ARN mutant et peut-être, selon le design de l'ASO, reconnue par les RNAses H qui dégradera le complexe ADN/ARN. L'ASO « (CAG)7 AON PS58 » apporté par électroporation ou injection dans des modèles murins DM1 réduit *via* l'action de la RNase H, entre 30-80% de l'expression de *DMPK* mutant (avec un effet tissu-dépendant). Les souris traitées présentent moins de *foci*, davantage de molécules MBNL1 libres, et certains épissages sont restaurés, mais la myotonie demeure présente (Mulders *et al.* 2009). Plus récemment, l'usage d'ASO LNA (Locked Nucleic Acids), n'activant pas la RNase H), fixent durablement les ARNs mutants dans des modèles cellulaires humains et dans un modèle *in vivo* de souris. Ces ASO LNA diminuent la présence de *foci*, empêchent la séquestration de MBNL1 par compétition et restaurent l'épissage d'ARNm sans changer la quantité de transcrits *DMPK* mutant dans le noyau (Wojtkowiak-Szlachcic *et al.* 2015).

D'autres stratégies utilisant **un dérivé Crispr-Cas9 visant l'ARN (dCas9)** ont également été testées en cellules COS-M6 : le complexe est porteur d'un guide permettant de fixer les ARNm *DMPK* mutants et est efficace pour les détruire, qu'il soit fusionné avec un domaine ARN nucléase (PIN) ou pas. Dans le premier cas, la fixation du Crispr provoque la dégradation de l'ARNm mutant par activité endonucléasique, dans le second cas (sans PIN), les ARNm mutants sont clivés de la même façon que lors d'usage d'ASO. Cette approche élimine également les *foci* dans des cellules de patients DM1, restaure la localisation de MBNL1 et rétabli l'épissage (Batra *et al.* 2017).

IV.3. THERAPIES GENIQUES

Ces dernières années, le développement d'outils d'édition du génome TALENs et Crispr/Cas9 a ouvert la voie à de nombreux essais d'édition du génome pour les maladies monogéniques. Dans le cas des maladies à TNRs, le but serait de réduire la taille des expansions à l'aide de ces outils. Différents essais en cellules humaines et dans des modèles murins ont déjà démontré l'attractivité de l'usage de Crispr/Cas9, dans les maladies à triplets, (voir pour exemples et revues (Richard 2015, Cinesi *et al.* 2016, Dastidar *et al.* 2018, Lee *et al.* 2018, Tremblay *et al.* 2019)) et a déjà brevetée dans certains cas (exemple : dans la FRDA,). Par exemple, dans le cas de la maladie de Huntington, le développement d'un complexe Crispr/Cas9 reposant sur une Cas9 mutante (Cas9 D10A ou Cas9 nickase) a fourni un outil capable de biaiser l'instabilité des triplet CAG vers les contractions dans un modèle de cellules humaines de la maladie de Huntington (Cinesi *et al.* 2016).

Dans la DM1, plusieurs études ont démontré l'intérêt de l'approche thérapeutique par Crispr/Cas9. Par exemple, il a été démontré dans des cellules de patients DM1 que l'utilisation de Crispr/Cas9 induisant une cassure de part et d'autre des répétitions CTG, ce qui élimine les longues répétitions de CTG. Ces coupures s'accompagnent d'une diminution du nombre de foci et restaure l'épissage alternatif, cependant des cas de off-target ont été détectés dans ces études, limitant ainsi pour l'instant leur utilisation chez l'Homme (Provenzano et al. 2017, van Agtmaal et al. 2017). Cette année, la première preuve de concept in vivo de l'efficacité de la thérapie par Crispr/Cas9 chez la souris a été publiée (Lo Scrudato et al. 2019). Dans cette étude, les auteurs se sont basés sur un système capable de cliver à l'intérieur des séguences flanguantes des répétitions CTG du locus DMPK humain. Cet outil est apporté par un vecteur rAAV. Ce système a été testé (1) dans les myoblastes de patients immortalisés et (2) par une unique injection intra-musculaire (dans le Tibialis Antérieur) chez des souris DM1 DMSXL âgées de 5 à 9 semaines, porteuses du locus DMPK humain mutant avec 1200 répétitions CTG. La zone coupée correspond à une séquence non codante, entre le codon stop et le signal de poly-adénylation en 3'UTR du gène DMPK. Aucune mutation « off-target » n'a été détectée. Les répétitions CTG sont efficacement coupées, bien que la région en aval des CTG soit plus résistante à la coupure peut-être à cause de structures secondaires rendant la zone moins accessible. Les souris traitées présentent moins de foci (-21 à -76%) mais l'expression de DMPK reste inchangée. De plus cette approche a également restauré le profil d'épissage de certains transcrits, comme MBNL1, BIN1, ATP2A1 notamment. La multiplication des preuves de concept, et d'outils plus précis pour traiter spécifiquement les allèles mutants ouvre à présent la voie à des essais thérapeutiques d'injection de Crispr/Cas9 chez les patients souffrant de pathologies à TNRs.

IV.4. CONCLUSION DE CHAPITRE : APERÇU DES ESSAIS CLINIQUES

Selon le NIH, il existe une quarantaine de programmes de recherche clinique sur la DM1 dans le monde. La DM1 étant une pathologie multisystémique, ces essais se concentrent souvent sur un des aspects de la pathologie, principalement par stratégies de replacement de médicament. Certains de ces essais ont déjà fourni des résultats dont nous proposons un bref résumé non exhaustif dans ce paragraphe.

Les atteintes musculaires, principal symptôme de la DM1, ont été le centre d'attention de nombreuses études. L'essai MYOMET en phase II a par exemple démontré un impact positif de la Metformine (antidiabétique) sur la mobilité des patients (Bassez *et al.* 2018). Différents essais se sont concentrés sur la myotonie : il a été mis en évidence que les facteurs bloquant le canal sodium comme la Lamotrigine (initialement développé pour l'épilepsie et les troubles de l'humeur) et le Mexiletine (régulateur du rythme cardiaque et traitement contre les douleurs chroniques), amélioraient significativement la myotonie. L'usage de la Lamotrigine

est aujourd'hui en phase III d'essai (essai n° NCT01939561) et la Mexiletine en phase II (Logigian *et al.* 2010).

D'autres essais se sont penchés sur les atteintes cognitives de la pathologie. Par exemple l'étude OPTIMISTIC repose sur le suivi d'un programme de thérapie cognitive et comportementale, et a récemment donné des résultats encourageants sur la fatigue sévère dont souffrent les patients DM1 (Okkersen *et al.* 2018). La phase II de test de l'utilisation dans le traitement de la DM1 de l'AMO-02, aussi appellé Tideglusib, (inhibiteur de GSK3β, initialement développé pour la maladie d'Alzeihmer et la médecine bucco-dentaire) montre une amélioration des fonctions cognitives chez des patients atteints de la forme congénitale et juvénile (essai n° NCT02858908).

Les résultats encourageant dans des modèles murins de l'utilisation d'ASO ont permis de lancer les premières études en phase clinique pour ce type d'approches. L'essai par la société IONIS, basé sur l'utilisation d'ASO DMPK-2.5RX visant à détruire les transcrits *DMPK* dans les muscles s'est arrêté en 2017 : la distribution de l'ASO n'était pas suffisante dans les tissus pour y corriger efficacement les défauts d'épissage. Cependant, cet assai a permis de mettre en avant la bonne tolérance des patients pour ce type de thérapies.

D'autres essais sont encore en cours aujourd'hui. Avec l'accumulation de preuves de concept *in vivo* qui s'accumulent pour la DM1 ou d'autres pathologies neuromusculaires et à triplets (voir (Babačić *et al.* 2019) pour revue), on peut raisonnablement supposer que les techniques d'édition du génome viendront compléter cet aperçu dans les années à venir (Cinesi *et al.* 2016, Monteys *et al.* 2017, Lo Scrudato *et al.* 2019). Une preuve supplémentaire de l'attrait de telles techniques pour la thérapie des maladies à TNRs est le dépôt récent du brevet n°US10323073B2 (Tremblay *et al.* 2019) pour l'usage de Crispr-Cas9 afin de stimuler l'expression de la Frataxine chez l'Homme.

V. OBJECTIFS DE LA THESE

L'instabilité des répétitions de microsatellites est un facteur clé pour les maladies à triplets instables, dont la DM1, maladie sujet de ce mémoire. La plupart du temps, ces pathologies présentent une accumulation du nombre de répétitions au cours de la vie des patients et de génération en génération en même temps que s'aggravent les symptômes. Cependant, il peut arriver qu'en de rares cas, des patients présentent des contractions intergénérationnelles (10% des transmissions paternelles et 3% des transmissions maternelles de la DM1). Des interruptions dans les répétitions (en moyenne 5% des cas de DM1, voir chapitre III.3.1) ont été reliées à certains cas de contractions intergénérationnelles et/ou à une mosaïque somatique réduite. Cependant, à ce jour, les mécanismes par lesquels les interruptions influenceraient l'instabilité des triplets restent souvent suggérés et peu connus.

L'objectif de cette thèse est de davantage comprendre l'implication des interruptions dans l'instabilité des triplets CTG chez les patients DM1. Pour cela, mon travail de thèse s'est appuyé sur 3 axes complémentaires.

V-1. FAMILLES DM1 ATYPIQUES ET INTERRUPTIONS

Avant le début de ma thèse, le laboratoire, en collaboration avec service diagnostic de l'hôpital Necker et le registre DM-Scope, a identifié deux familles DM1 atypiques, ne présentant pas de phénomène d'anticipation et une stabilisation voire des contractions de triplets CTG au cours de 3 à 4 transmissions maternelles successives. Au cours de mes premières années de doctorat, nous avons caractérisé ces 2 familles atypiques. Nous avons analysé différents facteurs connus pour influencer l'instabilité des TNRs comme la pureté des répétitions CTG, la transcription bidirectionnelle, le degré de méthylation. Par ailleurs, nous avons également comparé l'instabilité somatique de patientes de ces familles atypiques avec

celle de patients DM1 contrôles. Le résultat de ce travail a fait l'objet d'une publication dans Human Mutation en 2018 (Tome *et al.* 2018).



V-2. OPTIMISATION DU PROTOCOLE DE SMALL-POOL-PCR

La Small-Pool-PCR est l'expérience faisant consensus pour analyser l'instabilité des triplets répétés. Dans sa version en « Single-pool », elle permet d'amplifier une unique molécule d'ADN et d'apprécier l'instabilité des triplets à une très haute résolution. Cependant, les protocoles préexistants sont exigeants en matériels, demandent 3 jours incompressibles d'expérimentations avec de nombreuses manipulations notamment pour l'étape d'hybridation, sans contrôle possible au cours du protocole. Au cours de ma thèse, j'ai eu à réaliser un certain nombre de Small-Pool-PCR au cours desquelles j'ai trouvé pertinent d'optimiser cette méthode en termes de coûts, mais surtout en termes de temps et de nombre d'étapes nécessaires, limitant ainsi les biais expérimentaux. Le protocole résultant de ces optimisations a fait l'objet d'une publication dans le journal Biotechnique en 2018 (Dandelot *et al.* 2018).



V-3. ROLE(S) DES INTERRUPTIONS DANS L'INSTABILITE DES TRIPLETS CTG

La seconde partie de ma thèse a consisté à démontrer si les interruptions avaient un impact direct sur l'instabilité des triplets CTG, et, si oui, par quels mécanismes ? Pour cela, j'ai tout d'abord analysé *in vitro* l'impact d'une seule interruption CAG en 5' des répétitions CTG sur la structure secondaire de l'ADN ainsi que sur l'interaction des répétitions avec les protéines nucléaires. Ensuite, j'ai adapté un modèle cellulaire (Nakamori *et al.* 2010) reposant sur l'intégration aléatoire et stable de la région *DMPK* 3'UTR encadrée par des promoteurs permettant la transcription bidirectionnelle des répétitions CTG dans des cellules humaines. Nous avons ainsi analysé en parallèle des lignées cellulaires porteuses de 123 répétitions CTG pures ou interrompues. Avec ce modèle, j'ai pu étudier l'impact de l'interruption sur la transcription de *DMPK* mutant, la formation d'R-loop et l'instabilité somatique.

RESULTATS
I. FAMILLES DM1 ATYPIQUES ET INTERRUPTIONS

I.1. ÉLEMENTS DE CONTEXTE

Les interruptions ont été reliées à une stabilisation des triplets répétés dès 1994, avec les interruptions AGG identifiées dans les répétitions CGG sur le gène FMR1 responsables du syndrome de l'X Fragile (Eichler et al. 1994). Dès leur première description dans des familles DM1 en 2009 (Musova et al. 2009), les interruptions ont été associées à une instabilité atypique des triplets CTG: mosaïque somatique réduite ou contractions de triplets CTG. Le développement des techniques de séguençages et de TP-PCR (Triplet Primed PCR) ont permis de détecter plus précisément les interruptions. Ainsi, il a été estimé dans différentes études de cohortes de patients DM1 qu'entre 3 à 8% des patients DM1 étaient porteurs d'interruptions dans les triplets CTG (Musova et al. 2009, Braida et al. 2010, Santoro et al. 2013, Botta et al. 2017, Cumming et al. 2018). A travers ces études, différentes natures d'interruptions stables ou instables au cours des générations ont été identifiées dans les familles DM1 (CCG, CTC, CGG), localisés en 5', 3' ou au milieu des répétitions CTG (voir tableau 1 en Annexe). Les interruptions ont par ailleurs été associées à la stabilisation d'autres natures de TNRs comme par exemple les interruptions AGG et CAA stabilisant les répétitions CGG dans le FXS et la SCA2 respectivement (Eichler et al. 1994, Choudhry et al. 2001). Audelà de l'instabilité atypique des triplets répétés, les interruptions ont également été associées à des symptômes amoindris en comparaison avec des individus porteurs d'un nombre équivalent de répétitions pures, dans le cas de patients SCA1 et DM1 (Kraus-Perrotta et al. 2016, Santoro et al. 2017, Cumming et al. 2018). Ces observations ont conforté l'importance que peuvent avoir les interruptions dans la pathogénèse des maladies à TRNs. La multiplication des corrélations entre la présence d'interruptions et une instabilité atypique ainsi qu'une pathogénèse réduite a fait qu'aujourd'hui, les interruptions sont considérées comme des acteurs à part entière de l'instabilité des triplets répétés. Différentes études ont de plus montré que les interruptions influencent la méthylation, la formation de structures secondaires, ainsi que le niveau de transcription dans différentes pathologies à triplets (Pearson et al. 1998, Sakamoto et al. 2001, Mulvihill et al. 2005, Jarem et al. 2010, Santoro et al. 2015). Cependant les mécanismes par lesquels elles pourraient impacter les répétitions restent peu connus.

Dans l'article présenté ci-après, nous décrivons de nouvelles familles DM1 atypiques à l'instabilité intergénérationnelle biaisée vers les contractions et ne présentant aucun phénomène d'anticipation. Nous avons identifié des interruptions au sein des répétitions CTG de ces familles par TP-PCR et séquençage Sanger. Afin de davantage comprendre l'implication des interruptions dans l'instabilité des TNRs, nous avons pour la première fois

comparée par Small-Pool PCR la mosaïque somatique de ces familles à celle de patients DM1 contrôles. Nous avons par ailleurs analysé d'autres éléments connus pour moduler l'instabilité des triplets répétés tels que le statut de méthylation des séquences environnantes ou la transcription bidirectionnelle.

I.2. RESUME DES PRINCIPAUX RESULTATS

Nous décrivons ici le cas de 2 familles DM1 présentant 3 à 5 transmissions maternelles atypiques. En effet, ces familles ne présentent pas de phénomène d'anticipation et nous avons démontré par PCR que l'instabilité intergénérationnelle chez ces patientes était biaisée vers les contractions ou, du moins, était stabilisée.

I.2.1. IDENTIFICATION D'INTERRUPTIONS

Les interruptions étant connues pour être liées à une instabilité réduite des TNRs, voire des contractions, nous avons voulu déterminer si ces familles étaient porteuses d'interruption(s). Pour cela, nous avons réalisé une TP-PCR en 5' des répétitions. En effet, dans cet article, nous re-démontrons l'importance de ce type d'analyse en 5' et en 3' des répétitions CTG pour détecter au mieux les interruptions sur les extrémités des répétitions : la même expérience réalisée en 3' des répétitions ne permettait effectivement pas de détecter les interruptions, à l'instar de ce qui avait déjà pu être décrit dans les protocoles de diagnostic (Santoro *et al.* 2017). Par ailleurs, nous avons identifié par séquençage Sanger **un nouveau motif d'interruption : une unique interruption CAG en 5' des répétitions CTG**, transmise sur 5 transmissions maternelles. La seconde famille présente quant à elle plusieurs interruptions CCG en 5' des répétitions. Pour ces 2 familles, nous avons été en mesure de déterminer par combinaison de séquençage Sanger et de TP-PCR spécifique des interruptions, et que l'instabilité des triplets est prédominante en 3' des interruptions.

I.2.2. COMPARAISON DE LA MOSAÏQUE SOMATIQUE DES PATIENTES A INTERRUPTIONS AVEC CELLE DE PATIENTS DM1 CONTROLES

L'instabilité intergénérationnelle étant biaisée vers les contractions chez ces familles présentant des interruptions, nous avons voulu déterminer si l'instabilité somatique dans le sang de ces patientes était également atypique. Par Small-Pool PCR, nous avons comparé la mosaïque somatique de patientes des familles avec interruptions avec celle de patients DM1 contrôles dans le sang. Les contrôles DM1 ont des répétitions CTG pures, et ont été sélectionnés de telle sorte qu'ils aient le même âge au prélèvement et approximativement la même taille de triplets CTG héritée que les membres des familles atypiques analysés. Nous avons ainsi démontré que chez les familles porteuses d'interruption(s) CAG ou CCG, il y avait

toujours une mosaïque somatique biaisée vers les expansions, mais que l'amplitude des expansions était réduite.

Ainsi, l'instabilité somatique chez les familles à interruption(s) est biaisée vers les expansions, tandis que l'instabilité intergénérationnelle est biaisée vers les contractions. Ces résultats tendent à conclure que **les mécanismes régulant l'instabilité intergénérationnelle et somatiques seraient modulés différemment, ou sont peut-être différents.**

I.2.3. ANALYSE D'AUTRES FACTEURS MODULATEURS DE L'INSTABILITE DES TNRS

Il a été démontré que la transcription bidirectionnelle favorisait l'instabilité des triplets CTG (Nakamori *et al.* 2010) : nous avons donc testé la présence de transcrits sens et antisens de *DMPK* chez les patientes des familles atypiques. Nous avons détecté par RT-PCR l'expression sens et antisens de *DMPK* mutant dans des cellules PBL et LCLs de patientes de la famille A. Par ailleurs, comme il a été démontré que le statut de méthylation des séquences environnantes influençait l'instabilité des triplets (Dion *et al.* 2008) et que l'hyper-méthylation des régions en aval des répétitions CTG était notamment liée chez des patients CDM à la présence d'interruptions en 3' des répétitions (Santoro *et al.* 2015), nous avons analysé le statut de méthylation des séquences environnantes des triplets CTG dans les familles porteuses d'interruptions. Aucune différence dans le statut de méthylation n'a été détectée entre les patientes de familles atypiques et les patients DM1 contrôles avec des répétitions pures. Ainsi, le biais vers les contractions dans l'instabilité somatique ne peuvent pas être expliqués par (1) l'absence de transcription bidirectionnelle ou (2) un changement dans le statut de méthylation.

I.3. CONCLUSIONS GENERALE ET PERSPECTIVES

Dans cet article, nous présentons des familles DM1 très atypiques, notamment par l'absence de phénomène d'anticipation sur 3 et 4 générations successives à transmissions maternelles. En effet, dans le cas général, il aurait été attendu la survenue de la forme juvénile et/ou congénitale après plusieurs transmissions maternelles successives (Harper *et al.* 1992). Les différentes observations relayées ici suggèrent fortement un effet des interruptions sur l'instabilité somatique, dans la continuité des précédents travaux réalisés sur les interruptions dans la DM1 (voir (Santoro *et al.* 2017) pour revue). Par ailleurs, nous apportons ici des éléments laissant supposer **qu'une seule interruption CAG pourrait influencer l'instabilité des triplets CTG.** Cependant, à la seule lumière de cet article, il ne peut pas être complètement exclu que d'autres facteurs puissent être responsables ou co-responsables de l'instabilité atypique chez ces familles. Auquel cas les interruptions CTG.

I.4. CONTRIBUTION

Du point de vue expérimental, j'ai réalisé les expériences des figures 1B, 3 et S1 et S2 et certaines TP-PCR des figures 2(B,C). Ma participation à ces travaux a aussi consisté en la validation des contrôles DM1 (pureté et taille des interruptions par séquençage Sanger , PCR et TP-PCR), à l'analyse préliminaire chez la famille A et leurs contrôles DM1de l'instabilité des triplets (par Small-Pool PCR) et de l'expression de *DMPK* (par RT-qPCR). J'ai aussi activement participé à la conception et construction des figures 1, 2(B,C) et 3.

RESEARCH ARTICLE



Unusual association of a unique CAG interruption in 5' of DM1 CTG repeats with intergenerational contractions and low somatic mosaicism

Stéphanie Tomé¹ D | Elodie Dandelot¹* | Céline Dogan²* | Alexis Bertrand¹ | David Geneviève^{3,4} | Yann Péréon⁵ | DM contraction study group⁷ | Marie Simon³ | Jean-Paul Bonnefont³ | Guillaume Bassez⁶ | Geneviève Gourdon¹

¹Laboratory CTGDM, Inserm UMR1163, Paris, France; Institut Imagine, Université Paris-Descartes-Sorbonne Paris-Cité, Paris, France

²Neuromuscular Reference Center, AP-HP, Hôpital Pitié-Salpêtrière, F-75013, Paris, France

³Molecular Genetic Laboratory, Necker Hospital, Paris, France

⁴Département de Génétique Médicale, Maladies Rares et Médecine Personnalisée, CHU Montpellier, Université Montpellier, Montpellier, France

⁵Centre for Neuromuscular Diseases, Hôtel-Dieu Hospital, Nantes, France

⁶Sorbonne Université, Inserm, UMRS974, Neuromuscular Reference center, AP-HP, Hôpital Pitié-Salpêtrière, F-75013, Paris, France

⁷ Pauline Arnaud: Department of genetic, Bichat Hospital, Paris, France, Raphaële Chasserieau: Centre for Neuromuscular Diseases, Hôtel-Dieu Hospital, Nantes, France, Pascal Cintas: Neuromuscular Reference Center, Purpan Hospital, Toulouse, France, Ana-maria Cobo Esteban: Neuromuscular Reference Center, Marin Hospital, Hendaye, France, Marie-Carmen Cruz: Neuromuscular Reference Center, Purpan Hospital, Toulouse, France, Dalil Hamroun: Centre Hospitalo-Universitaire de Montpellier, Montpellier, France, Armelle Magot: Neuromuscular Reference Center, Hôtel-Dieu Hospital, Nantes, France, Alexandra Nadaj-Pakleza Neuromuscular Reference Center, Larrey Hospital, Angers, France, Anne-catherine Aube-Gauthier Neuromuscular Reference Center, Larrey Hospital, Angers, France, Andoni Urtizberea: Neuromuscular Reference Center, Marin Hospital, Hendaye, France

Correspondence

Tomé Stéphanie, Laboratory CTGDM, Inserm UMR1163, Paris 75015, France. Email: stephanie.tome@inserm.fr Gourdon Geneviève, Laboratory CTGDM, Inserm UMR1163, Paris 75015, France. Email: genevieve.gourdon@inserm.fr

*Dandelot Elodie and Dogan Céline contributed equally to this work.

Funding information

Contract grant sponsors: Institut National de la Santé et de la Recherche Médicale; Université Paris Descartes; AFM Telethon (16331 and 19757); National Research Agency (ANR-10-IAHU-01).

Communicated by Garry R. Cutting

Abstract

Myotonic dystrophy type 1 (DM1) is a dominant multisystemic disorder associated with high variability of symptoms and anticipation. DM1 is caused by an unstable CTG repeat expansion that usually increases in successive generations and tissues. DM1 family pedigrees have shown that ~90% and 10% of transmissions result in expansions and contractions of the CTG repeat, respectively. To date, the mechanisms of CTG repeat contraction remain poorly documented in DM1. In this report, we identified two new DM1 families with apparent contractions and no worsening of DM1 symptoms in two and three successive maternal transmissions. A new and unique CAG interruption was found in 5' of the CTG expansion in one family, whereas multiple 5' CCG interruptions were detected in the second family. We showed that these interruptions are associated with maternal intergenerational contractions and low somatic mosaicism in blood. By specific triplet-prime PCR, we observed that CTG repeat changes (contractions/expansions) occur preferentially in 3' of the interruptions for both families.

KEYWORDS

CTG contractions and 5' single CAG interruption, myotonic dystrophy type 1, triplet repeat instability

1 | INTRODUCTION

Several human neurological, neurodegenerative, and neuromuscular disorders are associated with unstable DNA microsatellite repeats including tri-tetra-penta-nucleotide tracts (Usdin, House, & Freudenreich, 2015). More recently, an unstable hexanucleotide repeat in the *C9ORF72* gene (MIM# 614260) was found associated to amy-

otrophic lateral sclerosis with frontotemporal dementia (ALS-FTD) (DeJesus-Hernandez et al., 2011; Renton et al., 2011). The first expanded repeats identified were the trinucleotide repeats (TNR) CAG/CTG and CGG/GCC that are associated, respectively, to spinal and bulbar muscular atrophy, Huntington disease (HD), some types of spinocerebellar ataxias (SCAs), myotonic dystrophy type 1 (DM1), and Fragile X syndrome (FXS) (Usdin et al., 2015). DM1 is a dominant

inherited disorder caused by one of the most unstable expanded CTG/CAG repeats, and it is located in the 3' untranslated region (3'UTR) of the DMPK gene (MIM# 605377) encoding a protein kinase (Brook et al., 1992; Mahadevan et al., 1992). CTG repeat expansions produce toxic CUG transcripts that accumulate in the nucleus, form RNA foci and disturb the function of nuclear proteins including MBNL and CELF proteins (Sicot & Gomes-Pereira, 2013). Clinically, DM1 is highly variable, shows a wide range of symptoms and different ages of onset. The adult form typically presents progressive distal muscle weakness, myotonia, cardio-respiratory problems but also ocular, endocrine, gastrointestinal, and cognitive dysfunction. The disease is progressive and often leads to significant disability. The congenital form of DM1 (CDM) is characterized by general hypotonia, respiratory distress, and feeding problems at birth and is often associated with delayed motor development and intellectual disability. Typically, CDM cases are associated with maternal transmissions (Harper, 2001).

In the general population, the CTG repeat length at the DM1 locus varies from 5 to 37 repeats and remains stable across generations and in tissues. Once above \geq 40 repeat tracts, the repeats become progressively more and more unstable. Intergenerational mutation rates may be as high as 100% per generation, with a striking tendency toward further repeat gains (about 90%) (Ashizawa & Harper, 2006). Intermediate alleles with 37-49 CTG tracts have been identified in unaffected parents of individuals with DM1 and have a tendency to expand (Martorell, Monckton, Sanchez, Lopez De Munain, & Baiget, 2001). Behavior of the CTG repeat between generations depends on repeat length and on the sex of the transmitting parent (Ashizawa, Dunne, Ward, Seltzer, & Richards, 1994; Brunner et al., 1993; Lavedan et al., 1993). Two large collaborative studies have estimated contraction frequencies at 10% in paternal transmissions and only 3%-4% in maternal transmissions (Ashizawa et al., 1994; Salehi et al., 2007). Typically, the CTG repeat length is positively correlated with symptoms severity and negatively correlated with age of onset, resulting in anticipation particularly obvious in DM1 (De Antonio et al., 2016; Harper, Harley, Reardon, & Shaw, 1992).

In DM1 families, inter-tissue variability of the CTG repeat length can be observed in fetuses at 13–16 weeks and the number of CTG repeats continue to expand throughout life after birth causing agedependent somatic mosaicism (Martorell et al., 1998; Martorell, Johnson, Boucher, & Baiget, 1997; Martorell, Martinez, Carey, Johnson, & Baiget, 1995; Wong, Ashizawa, Monckton, Caskey, & Richards, 1995). CTG repeat instability has been reported in various human tissues, including peripheral blood lymphocytes and muscles. In adult DM1 patients, the mutant allele size is typically larger in skeletal muscle than in peripheral blood leucocytes (PBLs) of the same individual (Anvret et al., 1993; Ashizawa, Dubel, & Harati, 1993; Thornton, Johnson, & Moxley, III, 1994; Zatz et al., 1995). The somatic mosaicism is particularly high in DM1 and progresses with age toward larger repeat lengths with different rates of expansion between tissues, possibly correlated with the progression of symptoms (Morales et al., 2012).

TNR instability is probably caused by the formation of unusual DNA triplet structures such as slipped-DNAs during different metabolic processes. Data from mouse models demonstrated that TNRs are

WILEY Human Mutation

unstable only when transcribed, suggesting that the transcription through TNRs is necessary to destabilize triplet repeats (Mangiarini et al., 1997). The formation of expansion is highly dependent on DNA repair, transcription, R-loop formation, DNA replication, and cis- elements such as chromatin packaging and both length and purity of TNR sequences (Polyzos & McMurray, 2017; Santoro, Masciullo, Silvestri, Novelli, & Botta, 2016; Usdin et al., 2015). Interrupted TNR tracts were first described in the 1990s in FXS (Eichler et al., 1994). In FXS, AGG interruptions were identified in the FMR1 allele (MIM# 309550) carrying less than 55 CGG repeats and the number of AGG interruptions is negatively correlated with the risk of full mutation supporting the role of interrupted sequences in the stabilization of TNR repeat loci (Eichler et al., 1994; Nolin et al., 2013; Yrigollen et al., 2014). CAT, CAA, and GAG interruptions are exclusively found in unexpanded alleles for SCA1, SCA2, and Friedreich ataxia (FRDA) and are associated with the stability of CAG and GAA unexpanded repeat tracts, respectively (Choudhry, Mukerji, Srivastava, Jain, & Brahmachari, 2001; Chung et al., 1993; Montermini et al., 1997). In FRDA, GGA interruptions in an expanded GAA sequence were also reported in one individual with late-onset ataxia (Ohshima et al., 1999). In SCA8, various interruptions such as a single CTT or multiple CCG and CTA tracts were found in the expanded CTG repeat alleles and were not clearly associated with a stabilization of CTG array (Hu et al., 2017; Moseley et al., 2000). In DM1, rare DMPK intermediate alleles revealed various CCG interruptions within CTG repeats (Braida et al., 2010; Leeflang & Arnheim, 1995; Musova et al., 2009). Interestingly, Leeflang and Arnheim (1995) observed that a DMPK allele with a (CTG)₄(CCGCTG)₁₆CTG pattern is more stable than an allele carrying 27 pure CTG repeats in single sperm analysis. Different DM1 cohort studies have revealed that repeat interruptions are observed in 3%-4% of French, 5% of Czech or 4%-8% of Italian DM1 patients (Botta et al., 2017; Braida et al., 2010; Musova et al., 2009; Santoro et al., 2013). CCG, GGC, and CTC interruptions have been first reported at the 3' end of the CTG repeat expansion in DM1 patients (Botta et al., 2017; Braida et al., 2010; Musova et al., 2009; Pesovic et al., 2017; Santoro et al., 2013). Recently, multiple CCG interruptions in the 5' end of the CTG repeat sequence were described in four DM1 patients from two distinct DM1 families. These results show that CCG interruptions are not exclusively located in the 3'end of the CTG repeat expansion (Botta et al., 2017; Dryland, Doherty, Love, & Love, 2013). All these interruptions were associated with stabilization or contraction of the expanded CTG repeat tracts through maternal and paternal transmissions (Botta et al., 2017; Braida et al., 2010; Musova et al., 2009). In addition, using mathematical models, Braida et al. (2010) showed that the degree of somatic mosaicism is low in an atypical DM1/Charcot Dutch family compared with the expected somatic mosaicism for the corresponding age of sampling and progenitor allele length (Higham & Monckton, 2013). Different studies have shown that formation of secondary DNA structures, nucleosome positioning, methylation status and transcription level can be affected by the AGG, CAT, GGA, and CCG interruptions described in SCA1, FXS, FRDA, and DM1 (Jarem, Huckaby, & Delaney, 2010; Mulvihill, Nichol Edamura, Hagerman, Pearson, & Wang, 2005; Pearson et al., 1998; Sakamoto et al., 2001; Santoro et al., 2015; Volle & Delaney, 2013). However, the role of interruptions in the

971

WILEY Human Mutation

mechanisms of TNR stabilization/contraction remains elusive, especially for DM1.

The aim of our study is to gain additional insights on the mechanisms of CTG repeat contractions observed in some DM1 families. We therefore characterized the sequences and dynamics of CTG repeat tracts across generations. For the first time, we studied the somatic mosaicism in blood from two unusual DM1 families that exhibit apparent intergenerational contractions of the CTG repeat compared with DM1 patients carrying pure repeat and matched in age at sampling and in inherited CTG repeat length. Our data suggest that a single nucleotide change in the 5' end of the CTG repeat array is sufficient to prevent expansions across generations and to decrease somatic mosaicism.

2 | MATERIALS AND METHODS

2.1 DM1 samples

DM1 patients from families A and B were recruited by the Genetics department of the Necker-Enfants Malades Hospital and DM-Scope registry. DM1 patients with pure repeats were selected from DMscope registry (Dogan, Puymirat, & Bassez, 2015). The study was performed in collaboration with the Genethon Biobank (BACG) which has been approved by the French data protection authority (National Commission on Informatics and Liberty [CNIL] and the French Ministry of research, N° AC-2007-18, (06/25/2008). Blood samples were collected according the following consent procedure that was approved by the national ethic committee CCTIRS (Advisory Committee on Information Processing in Material Research in the Field of Health): adult patients or the legal guardians received an information letter presenting the study. A written informed consent was required before collecting blood samples and sending them to the Genethon Biobank. Post mortem heart of DM1 fetus was sampled at autopsy twenty years ago and no law required approvals at the time of this sample collection (Michel, Huguet-Lachon, & Gourdon, 2015). The length of the CTG repeats, the age of sampling and symptoms are listed in the Table 1 and Supp. Table S1 for each individual. Lymphoblastoid cell lines (LCLs) were available from the A.3, A4.1, and A4.2 individuals. PBLs from the A4.1 and A4.2 individuals were also available.

2.2 | Estimation of inherited CTG repeat length in DM1 patients

To estimate the inherited CTG repeat length in DM1 patients, 5 ng of DNA from blood or trophoblast was amplified in a 25 μ l reaction using 0.4 μ M ST300F (5'-GAACTGTCTTCGACTCCGGG-3') and ST300R (5'-GCACTTTGCGAACCAACGAT-3') primers, 1 × Custom master mix (Thermo Fischer Scientific, Vilnius, Lithuania) and 0.04U Thermoperfect *Taq* polymerase (Peak International Product b.v, Eerbeek, The Netherlands). The following cycling conditions were used: 5 min at 96°C; 45 sec at 96°C, 30 sec at 60°C and 3 min at 72°C (30 cycles); 1 min at 60°C and 10 min at 72°C (1 cycle). PCR product was mixed

with DNA-loading dye and subjected to electrophoresis on a 1.5% agarose gel at 120 V for 1–2 hr. The size of the PCR products was measured using Bio-Rad's Image Lab software. The approximate number of triplet repeats can be obtained by subtracting 361 bp (corresponding to the size of 5' and 3' flanking regions) to the PCR product length divided by 3. To complete our analysis, PCR was performed with 101 (5'- CTTCCCAGGCCTGCAGTTTGCCCATC-3') and 102 (5'-GAACGGGGCTCGAAGGGTCCT GTAGC-3') primers (Seznec et al., 2000). The approximate number of triplet repeats can be obtained by subtracting 112 bp (corresponding to the size of 5' and 3' flanking regions) to the PCR product length divided by 3.

2.3 5' Triplet prime PCR

To analyze the purity of the CTG repeat tract at the 5' end of the CTG repeat array, triplet prime PCR (TP-PCR) was performed on both strands of the CTG repeat (Warner et al., 1996). 5' of the CTG repeat was amplified using the primer upstream of the CTG repeat DM-C-FAM (5'FAM-AACGGGGCTCGAAGGGTCCT-3') and P3R (5'-TACGCATCCCAGTTTGAGACG-3') and P4CAG (5-'TAC GCATCCCAGTTTGAGACGCAGCAGCAGCAGCAGCA-3') primers (Warner et al., 1996). 20-100 ng of DNA from blood or trophoblast was amplified in a 25 μ l reaction using 0.4 μ M of FAM and P3R primers and 0.04 μ M of P4CAG primer, 1 \times Custom master mix (Thermo Fischer Scientific, Vilnius, Lithuania) and 0.06 U Thermoperfect Tag polymerase (Peak International Product b.v, Eerbeek, The Netherlands). The following cycling conditions were used: 5 min at 94°C; 60 sec at 94°C, 1 min 30 sec at 68°C and 2 min at 72°C (30 cycles) and 10 min at 72°C (1 cycle). The amplified product (1–3 μ l) was analyzed using 3500XL genetic analyzer (Applied Biosystems, Foster city, CA, USA) and Gene Mapper software. GeneScan 600 LIZ dye size standard v2.0 is used in this experiment (Thermo Fisher Scientific, Courtaboeuf, France). The method analyzing the purity of the CTG repeat tract at the 3' end of the CTG repeat array is detailed in Supplementary Material.

2.4 | Sequencing

PCR using ST300F and ST300R was performed to sequence normal and expanded CTG repeat tracts. PCR products were purified using NucleoSpin Gel and PCR Clean-Up according to manufacturer recommendation (Macherey Nagels, Hoerdt, France). PCR products were sequenced either by direct sequencing using primer ST300F and the BigDye Terminator v3.1 cycle sequencing kit (Thermo Fischer Scientific, Courtaboeuf, France) or cloning approach. Purified PCR products were cloned using TOPO-TA cloning kit (Thermo Fischer Scientific, Courtaboeuf, France) and each clone was sequenced using primers M13F, M13R and BigDye Terminator v3.1 cycle sequencing kit. Direct sequencing PCR reactions involved 1 cycle of denaturation 96°C (5 min) and 30 cycles of 96°C (30 sec), 58°C (15 sec), 60°C (45 sec). Sequence reaction products were sequenced using 3500xL Genetic analyzer (Applied Biosystem, Foster city, CA, USA) and analyzed using Macvector software.

TOMÉ ET AL.

973

				Interruption				DM1 features	
Sample	Sex	Normal allele length	Expanded allele length (nCTG)	5'	Middle	3′	Interruption pattern (5'-3')	Age of consultation and sampling (yo)	DM1 symptoms
A1	F	21	170	Yes	No	No	(CTG)31(CAG)1(CTG)x-170	69	Severe muscular weakness -Feeding difficulties - Nasogastric enteral feeding 72 yo: death
A2	F	5	150	Yes	No	No	(CTG)31(CAG)1(CTG)x-150	49	Severe muscular weakness – Myotonia -Heart conductive defect - Implantable cardiovector defrillator 58 yo: Heart attack
A3	F	5	140	Yes	No	No	(CTG)29(CAG)1(CTG)x-140	32	Asymptomatic
A4.1	F	5	125	Yes	No	No	(CTG)31(CAG)1(CTG)x-125	13	Asymptomatic
A4.2	F	21	130	Yes	No	No	(CTG)31(CAG)1(CTG)x-130	7	Asymptomatic
A4.3	-	21	125	Yes	No	No	(CTG)30(CAG)1(CTG)x-125	-	Therapeutic abortion
B1	F	13	365	Yes	ND	No	(CTG)11(CCG)1(CTG)2(CCG)1 (CTG)4(CGG)1(CTG)n	58	Severe muscular weakness - Feeding difficulties - Nasogastric enteral feeding
B2	F	11	310	Yes	ND	No	(CTG)11(CCG)1(CTG)2(CCG)1 (CTG)4(CGG)1(CTG)n	33	No muscular weakness-Myotonia
B3.1	-	13	300	Yes	ND	No	(CTG)11(CCG)1(CTG)2(CCG)1 (CTG)4(CGG)1(CTG)n	-	Therapeutic abortion
B3.2	-	13	235	Yes	ND	No	(CTG)11(CCG)1(CTG)2(CCG)1 (CTG)4(CGG)1(CTG)n	-	Therapeutic abortion
B3.3	-	5	250	Yes	ND	No	(CTG)11(CCG)1(CTG)2(CCG)1 (CTG)4(CGG)1(CTG)n	-	Therapeutic abortion

TABLE 1 Clinical and molecular data in DM1 patients showing apparent contractions over maternal transmissions

2.5 | Specific TP-PCR

2.6 | Small pool PCR

Small pool PCR (SP-PCR) amplifications and PCR product electrophoretic analyses were performed using the methods previously described (Gomes-Pereira, Bidichandani, & Monckton, 2004). Blood DNA samples were digested with *Hin*dIII enzyme and SP-PCR was performed using DM-C and DM-BR primers (Monckton, Wong, Ashizawa, & Caskey, 1995). The DNA was denatured by heating to 96°C (5 min). PCR reactions involved 28 cycles of 96°C (45 sec), 68°C (45 sec), and 70°C (3 min) with a chase of 1 min at 68°C and 10 min at 70°C. SP-PCR products were loaded on the same 1,5% agarose gel to compare instability between DM1 patients from families A and B and their DM1 respective matched DM1 patients and ran at 160 V for 16–20 hr. The PCR products were then transferred to GeneScreen Plus Hybridization transfert Membrane (Perkin Elmer, Villebon sur Yvette, France) and detected by hybridization using a nonradioactive method as described in Tome, Nicole, Gomes-Pereira, and Gourdon (2014).

2.7 | RNA isolation and reverse transcription

RNAs were extracted from LCLs and PBLs cells as described in Huguet et al. (2012). Briefly, cells were homogenized in Trizol (Thermo Fischer Scientific, Courtaboeuf, France) using a tissue lyser and total RNA was extracted using PureLink RNA miniKit (Ambion, Thermo Fischer Scientific, Courtaboeuf, France) according to the manufacturer's protocol. To analyze the expression of *DMPK* transcripts, cDNA was synthesized using Superscript II reverse transcriptase (Thermo Fischer Scientific, Courtaboeuf, France) and random hexamer primers from 1ug of RNA from LCLs. To investigate the expression of antisense or sense *DMPK* transcripts, cDNA was synthesized using Superscript III reverse transcriptase (Thermo Fischer Scientific bought Lifectechnologies) and strand specific primers (Primer sens LK-RTDMPK3S 5'-CGACTGGAGCACGAGGACACTGACGCCTGCCAGTTCACAACCGCT CCGAGCGT-3' and primer antisens LK-AMH 5'-CGACTGGAGCAC

WILEY Human Mutation

GAGGACACTGACGCCTGCCAGTTCACAACCGCTCCGAGCGT-3') from 1 μ g of RNA from LCLs and from 125 ng from PBLs according to manufacturer's protocol. RNAs were tested with retrotranscriptase (RT+) and without retrotranscriptase (RT-) to check DNA genomic contamination.

2.8 | PCR

The expression of the DMPK transcripts was analyzed using DM-C and DM-BR primers. The expression of sense DMPK transcripts was analyzed using LK3 (5'-GGAGCACGAGGACACTGA-3') and 102 (5'-GAACGGGGCTCGAAGGGTCCTTGTAGC-3') primers. cDNAs were amplified in a 25 μ l reaction using 0.4 μ M of primers, 1 × buffer, 0.5 mM MgCl2, 10% Glycerol, 0.5 mM dNTP and 0.04U Thermoperfect Taq polymerase (Peak International Product b.v, Eerbeek, The Netherlands). The following cycling conditions were used: 5 min at 96°C; 45 sec at 96°C, 30 sec at 68°C and 3 min at 72°C (30 cycles); 1 min at 68°C and 10 min at 72°C (1 cycle). The expression of antisense DMPK transcripts was analyzed using LK3 and antiN3H3 (5'-TGCGAACCAACGATAG-3') primers. cDNAs were amplified in 25 μ I reaction using 0.4 μ M of primers, 1 × buffer, 1 mM MgCl₂, 10% glycerol, 0.5 mM dNTP and 0.04U Thermoperfect Taq polymerase (Peak International Product b.v, Eerbeek, The Netherlands). The following cycling conditions were used: 5 min at 96°C; 30 sec at 96°C, 30 sec at 60°C and 1 min at 72°C (30 cycles); 1 min at 60°C and 10 min at 72°C (1 cycle).

2.9 | Bisulfite-sequencing PCR

DNA (50-200 ng) from blood of DM1 patients was bisulfite treated using the Imprint DNA modification Kit (Sigma-Aldrich, Lyon, France) according to one step modification procedure. Bisulfite reaction time is adapted for DNA template contains high GC content or secondary structure (150 min). Bisulfite-treated DNA was quantified by Qubit fluoremetric quantification (Thermo Fischer Scientific, Courtaboeuf, France) and amplified as described in Barbe et al. (2017). Upstream and downstream of the CTG repeat tracts (CTCF1 and CTCF2 regions, respectively) were amplified by nested and hemi-nested PCR (Barbe et al., 2017). For the first PCR, 2.5-5 ng of bisulfite-treated DNA was amplified in a 25 μ l reaction using 0.4 μ M primers, 1 × Custom master mix (Thermo Fischer Scientific, Courtaboeuf, France) and 0.04U Thermoperfect Taq polymerase (Peak International Product b.v, Eerbeek, The Netherlands). The following cycling conditions were used: 5 min at 94°C; 30 sec at 94°C, 30 sec at 55°C and 30 sec at 72°C (40 cycles); 5 min at 72°C (1 cycle). PCR products (3–6 μ l) from the first round were amplified following the same conditions of the first PCR. Amplicons were resolved by running 2% agarose gel electrophopheresis. The length of PCR products obtained from CTCF1 and CTCF2 regions is 294 and 169 bp, respectively. PCR products were purified using NucleoSpin Gel and PCR Clean-Up according to manufacturer recommendation (Macherey Nagels, Hoerdt, France). Purified PCR products were cloned using TOPO-TA cloning kit (Thermo Fischer Scientific, Courtaboeuf, France) and each clone (15-60 ng) was sequenced using primers M13R and BigDye Terminator v3.1 cycle sequencing kit. Sequence reaction products were sequenced using 3500xL Genetic analyzer (Applied Biosystem, Foster city, CA, USA) and analyzed using quantification tool for methylation analysis (QUMA). Samples showing less that 95% of bisulfite conversion and more that 10% of DNA mismatches were excluded of our analyses. At least 20 clones sequencing were performed to quantify the level of the methylation of each CpG site.

3 | RESULTS

3.1 | Apparent CTG repeat contractions in successive maternal transmissions

Among 51 French DM1 families followed in the Genetics department of the Necker-Enfants Malades Hospital (corresponding to about 150 transmissions), two families (families A and B) showed maternal intergenerational CTG repeat contractions (Figure 1). PCR amplification of the CTG repeat array using blood DNA revealed a decrease of CTG repeat length across three generations and five maternal transmissions in family A (Figure 1B). In family B, no apparent intergenerational expansion was observed in four maternal transmissions across two successive generations (Figure 1). In addition, despite adult symptoms reported in the mothers A1, A2, and B1 from the first generations in both families, no congenital, juvenile or childhood forms of DM1 were observed after two to three generations and four maternal transmissions suggesting no worsening of DM1 symptoms across generation in these two families (Table 1).

3.2 | Identification of a unique CAG in the 5' end of the CTG repeats in family A

In order to characterize the nature of the expanded CTG repeat sequences in family A, we performed TP-PCR both at the 5' and 3' ends of the CTG repeat, on blood DNA from different members of this family (Figure 2B and Supp. Figure S1A). By this technique, we covered all CTG repeat expansions in DM1 members that carry less than 150 CTG repeat tracts. For all DM1 patients, a distinct gap is observed in the contiguous peak pattern of the 5' TP-PCR products whereas the peak pattern of the 3' TP-PCR products is regular (Figure 2B and Supp. Figure S1A). The peak pattern revealed the presence of interruption(s) at the 5' end of the CTG repeat expansion in all DM1 members. The 3' end and the middle part of the CTG repeat sequence showed no interruption (Supp. Figure S1A). Surprisingly, we did not detect a clear gap using 3' TP-PCR that also covers the full CTG repeat tracts, probably due to lower fluorescent signal at the end of the 3' TP-PCR products. In order to identify the 5' interruption(s) in this family, we combined cloning and direct sequencing of the CTG repeat expansions amplified by PCR using two primers located upstream and downstream of the repeat. For the first time in DM1 families, we identified a unique CAG triplet closer to the 5' end of the CTG repeat (Figure 3A). Digestions of the CTG repeat PCR products with Alul enzyme (that recognizes AGCT motif) confirmed the presence of a CAG repeat interruption (Supp. Figure S2). By haplotype analysis at the DM locus in members of family A, we observed that this interrupted CTG repeat expansion is associated with the common DM1 haplotype (Haplotype

974



FIGURE 1 Atypical DM1 families A and B. A: Part of families A and B pedigrees (Standardized human pedigree nomenclature according to Bennet et al., 2008) (Bennett, French, Resta, & Doyle, 2008). Individuals with crosshatches represent asymptomatic/presymptomatic carrier. Approximate CTG expansion number is placed in parentheses. B: PCR amplification of the *DMPK* alleles. Normal and expanded alleles are indicated in bracket. The scale on the left represents the positions and sizes of the molecular weight markers in bp (M100 and M1kb from Thermo Scientific). Well B is the blank



FIGURE 2 A: On the top: Schematic diagram depicting the location of repeat-primed PCR primers used to amplify either the 5' or 3' end of the CTG repeat expansion. It should be noted that annealing of the repeat-primed primers can also overlap. On the bottom: A typical trace of a DM1 patient with pure repeats. **B**: 5'TP-PCR at the DMPK CTG repeat locus in family A. **C**: 5'TP-PCR at the DMPK CTG repeat locus in family B. 3' TP-PCR at the DMPK CTG repeat locus is representing in supplementary material. The y-axis represents the intensity of fluorescence (arbitrary unit) whereas the x-axis represents the size in bp of each PCR product. Each fragment obtained by TP-PCR is represented by one peak



FIGURE 3 Sequence results containing the 5' interruption motif. **A**: A3 individual from family A (sequence obtained by cloning approach). **B**: B3.2 fetus from family B (sequence obtained by direct PCR product sequencing)

A) as described in Neville, Mahadevan, Barcelo, and Korneluk (1994) (data not shown).

3.3 | Detection of three nonconsecutive CCG interruptions in the 5' end of the CTG repeat in family B

In family B, carrying more than 235 triplet repeats, bidirectional TP-PCR revealed the presence of interruptions at the 5' end of the CTG repeat expansion in five members (four transmissions and two generations; Figure 2C). In contrast, the 3' end of the repeat appeared free of interruption (Supp. Figure S1C). The TP-PCR method did not allow the characterization of the middle of the expanded repeats in this family. By cloning and sequencing the CTG repeat tracts, we identified three nonconsecutive CCG interruptions in the 5' of the expansion. All the DM1 members in this family carry a new identical motif $(CTG)_{11}(CCG)_1(CTG)_2(CCG)_1(CTG)_4(CCG)_1$ in the 5' end of the tracts that have not been described before (Figure 3B). The presence of CCG interruptions was confirmed by digestion using *Aci*l enzyme (Supp. Figure S3).

3.4 | CTG repeat instability predominantly occurs in 3' of the interruptions

In order to localize directly the interruptions in the CTG repeats from both families A and B, we combined direct sequencing and specific 5' TP-PCR with one primer covering the interruption and a second primer located upstream of the CTG repeat array (Figure 4 and data not shown). By specific TP-PCR, we observed a major peak corresponding to the interrupted DMPK allele products in family A (Figure 4A). DM1 individuals A1 and A2 carry the CAG repeat interruption at the 32nd trinucleotide position, A3, A4.1, and A4.2 at the 30th position and A4.3 at the 31st position. This observation suggests that small contractions and expansions (-2 or +1) can occur in 5' of the interruption but the majority of the repeat length difference between individuals arose in 3' of the interruptions. In family B, as expected, 5' specific TP-PCR showed three distinct peaks corresponding to three nonconsecutive CCG repeat interruptions (Figure 4B). The motif (CTG)₁₁(CCG)₁(CTG)₂(CCG)₁(CTG)₄(CCG)₁ was found at the same position in the five DM1 individuals indicating that the CTG repeat length changes between individuals occurred in 3' of the interruptions.

3.5 | 5' Interruptions decrease the level of somatic mosaicism in blood

In order to analyze the CTG repeat somatic mosaicism in blood, SP PCR was performed using blood DNA from A2, A3, and B2 individuals



FIGURE 4 Specific 5' TP-PCR at the DMPK CTG repeat locus in families A and B. The y-axis represents the intensity of fluorescence (arbitrary unit) whereas the x-axis represents the size in bp of each PCR product. The minor peaks observed in the electropherograms can be explained by the somatic instability and/or polymerase slippage during the CTG repeat amplification



FIGURE 5 Somatic mosaicism in the A2, A3, and B2 individuals compared with the DM1 patients carrying pure repeat and matched in age and CTG repeat length. The approximate CTG expansion number is enclosed in parentheses. The autoradiographs show representative SP-PCR analyses of DNA from blood of each DM1 patient. The sizes of DNA molecular weight marker III, DIG-labeled are indicated on the left of the figure. About 5–10 DNA molecules were amplified for the A2, A3 members and DM1 patients with pure repeats. About 3–5 DNA molecules were amplified for B2 and DM1 patients with pure repeats.

(Figure 5). DM1 patients with pure CTG repeat expansions and matched in age and CTG repeat lengths as close as possible were used for comparison (Supp. Table S1). The extent of somatic mosaicisms appeared lower for A2, A3, and B2 individuals compared with the respective matched DM1 patients with pure repeats, despite the heterogeneity observed between these DM1 patients (Figure 5). These results indicate that 5' interruptions are associated with a lower rate of somatic instability. However, CTG repeat expansions can still be observed for A2, A3 and B2 individuals suggesting that the CTG repeat length could increase over time (Figure 5). SP-PCR experiments performed at two time points over a period of about 10 years revealed no detectable increase of the somatic mosaicism in A3 (Figure 6A). However, individual B2 showed a small shift of the mosaicism toward expansions with age (Figure 6B). Classical PCR amplifications with the different ages at sampling support the apparent intergenerational contractions in family A and at least stabilization of the repeat over maternal transmission in family B (Figure 6C and D).

3.6 | Expression of DMPK in family A

In order to understand how a single nucleotide change in 5' can induce the unusual behavior of the CTG repeat in family A, we checked if *DMPK* gene is expressed or not in both direction in available samples from A3, A4.1, and A4.2 by RT-PCR. We observed that normal and expanded *DMPK* alleles are transcribed in PBLs and in LCLs (Figure 7A). Using specific primers, we observed that antisense transcripts for the normal and expanded *DMPK* alleles are expressed in LCLs (Figure 7B). An absence of expanded *DMPK* transcripts cannot explain the atypical dynamic of the CTG repeats observed in family A.

3.7 | No differential CpG methylation profile in families A and B

Recently, it was reported that multiple CCG interruptions at the 3' end of CTG repeat tracts are associated with a hypermethylation of the region downstream of the CTG repeat expansions. To investigate

if a unique CAG or multiple CCG interruptions, located at the 5'end of repeat tracts, could also alter the levels of methylation in the flanking regions, we used bisulfite sequencing. We found no difference in the CpG methylation status between A3 and B2 individuals and matched DM1 patients with pure repeats (Figure 8).

4 | DISCUSSION

In this study, we report two unusual French DM1 families showing apparent CTG repeat contractions over three and two maternal successive transmissions corresponding to five and four maternal transmissions in each family, respectively. No juvenile or congenital forms were observed in these two families, although it would have been expected that both forms of DM1 would be transmitted after several successive maternal transmissions (Harper et al., 1992). These observations suggest that one or more genetic factor(s) prevent CTG repeat expansions across maternal transmissions in these families. We describe, for the first time, a unique CAG interruption in 5' of the expanded CTG repeat in one family and, as previously described in Italian families, three nonconsecutive CCG interruptions (Botta et al., 2017). However, the location and number of CCG interruptions differ between the French and Italian families. In family A, the CAG interruption is associated with the DM haplotype suggesting that pure and interrupted repeat expansions shared the ancestral haplotype as observed in SCA2 (Ramos et al., 2010). This observation suggests that the CAG interruption has arisen as de novo base substitution in this family. This hypothesis is supported by the direct observation of de novo CCG or CTC interruptions across one or several generations in DM1 families (Botta et al., 2017; Braida et al., 2010; Musova et al., 2009; Pesovic et al., 2017). Identification of a unique CAG interruption in one of our French DM1 families with apparent intergenerational contractions suggests that not only multiple interruptions but also unique interruption are sufficient to prevent CTG repeat expansions through several maternal transmissions. We also found that the extent of somatic mosaicism decreases in blood from patients carrying the



978



FIGURE 6 A and B: Somatic mosaicism in the A3 and B2 individuals at two ages of sampling. The sizes of DNA molecular weight marker III, DIGlabeled are indicated on the left of the figure. About 2–3 DNA molecules were amplified for the A3 individual and about 5–15 DNA molecules were amplified for B2 patient. **C**: PCR amplification of *DMPK* alleles in family A. **D**: PCR amplification of *DMPK* alleles in family B. Age of sampling is placed in parentheses. The upper band corresponds to expanded allele and lower band to normal allele. The scale on the left represents the positions and sizes of the molecular weight markers in bp (M100 and M1kb+ from Thermo Scientific). Well B is the blank

5' unique CAG interruption or the 5' multiple CCG interruptions. Although somatic mosaicism (frequency of expansions or CTG repeat size changes) is altered in these DM1 patients, no apparent contraction was observed in blood. Moreover, B2 patient clearly showed somatic expansions with age. These findings suggest that the behavior of interrupted CTG repeat expansion differs between germline and soma. TNR instability is sensitive to the level of DNA repair proteins, replication profiles, and transcription (Cleary & Pearson, 2005; Jones, Houlden, & Tabrizi, 2017; Pearson, Nichol Edamura, & Cleary, 2005; Santoro et al., 2016; Tome et al., 2013; Tome et al., 2011). The efficiency of the different processes involved in TNR instability probably varies in somatic and germinal tissues and could be affected by the presence of interruptions in CTG repeat expansions in different ways.

Interruptions in CTG repeat expansions allowed analysis of the polarity of the CTG repeat tract changes in the two DM1 French families using specific TP-PCR. Small CTG expansions and contractions were observed in 5' of the unique CAG interruption whereas no change was detected in 5' of the multiple CCG interruptions. Despite few CTG repeat changes in 5' of the CAG interruption, CTG repeat changes preferentially occurred in 3' of the interruptions in both families A and B. Our data indicate that CTG repeat changes (contractions and/or expansions) are polar in DM1 patients carrying interrupted CTG repeat expansion. Interestingly, the 3' polarity of the tract length change was also observed in yeast carrying a CAG repeat tract interrupted by a unique CAT interruption (Maurer, O'Callaghan, & Livingston, 1998).

Little is known about the mode of action of interruptions on the dynamics of TNRs instability. Studies in different model systems have shown that transcription through CTG/CAG repeats is involved in TNR instability (Dion, 2014). Mangiarini et al. (1997) observed in HD mouse models that expanded CAG repeats are unstable only in the lines expressing the huntingtine transgene. These data support the hypothesis according which transcription of TNR loci is necessary to destabilize the CTG/CAG repeat tracts. More recently, Nakamori, Pearson, and Thornton (2011) have demonstrated that bidirectional transcription of DMPK further destabilizes CTG repeat expansions in transfected human cells in comparison with unidirectional transcription. In our study, we observed that normal and mutated alleles are both

TOMÉ ET AL.



FIGURE 7 A: Expression of sense DMPK transcripts in PBLs and LCLs. B: Expression of antisense DMPK transcripts in LCLs. Normal and expanded alleles are indicated in bracket. The extra bands are nonspecific bands generated during the PCR amplification. The sizes of the molecular weight marker are indicated on the left in bp (M100 from Thermo Scientific)



FIGURE 8 A: DMPK regions analyzed by bisulfite sequencing PCR with the inclusion of CTCF1 and CTCF2 binding sites represented in gray shadows, respectively. B. CpG methylation levels from the A3, A2 individuals and their respective matched DM1 patients. Top panel: Heart from DM1 fetus: 16 week of gestation (>1,500 CTG). White dots, 0% of methylation; Black dots 100% of methylation

transcribed bidirectionally in LCLs and/or PBLs from family A. Therefore, these data suggest that the absence of expanded DMPK transcripts cannot explain the atypical dynamic of the CTG repeats observed in this family. Previous studies have shown that CpG methylation near expanded CTG/CAG repeats change the dynamics of repeat expansions in SCA1 and SCA7 mouse models (Dion, Lin, Hubert, Waterland, & Wilson, 2008; Libby et al., 2008). In addition, Santoro and his colleagues have found that the 3'CCG interrupted expanded alleles were associated with CpG hypermethylation downstream of the CTG repeat expansion compared with pure expanded alleles in patients carrying more than 400 CTG repeats (Santoro et al., 2015). However, we found no difference in the levels of CpG methylation between patients carrying interrupted or pure CTG repeat expansions. These results suggest that methylation in the vicinity of the repeat tract is not the major mechanism involved in the unusual instability observed in families A and B. The differences in the CpG methylation profiles between our two families and the families reported by Santoro et al. could be explained by differences in location, type, and pattern of interruptions and inherited repeat lengths.

In vitro analyses of interrupted TNR showed that interruptions limit the formation of secondary structures formed by TNR in SCA1, FRAXA, and FRDA (Jarem et al., 2010; Pearson et al., 1998; Sakamoto et al., 2001). Pearson et al. (1998) have observed that a unique AGG interruption in the expanded CGG repeat tract is sufficient to weaken the formation of slipped strand DNA structure. Thus, a single CAG interruption in the 5' end of the CTG repeat expansions could modulate the formation and/or the stability of hairpin structures that could impact the dynamics of instability by altering DNA repair or the replication fork progression. Destabilization of interrupted CTG repeats has been observed in MMR-deficient yeast suggesting that mismatch repair proteins such as MSH2 are important to stabilize the interrupted CTG sequence in this model (Rolfsmeier, Dixon, & Lahue, 2000). We can assume that the mismatch repair proteins may also stabilize the interrupted CTG repeat in humans.

⁹⁸⁰ WILEY Human Mutation

In conclusion, our data describe DM1 patients with a new pattern of CCG interruptions at the 5' end of the CTG repeat expansion as well as a unique CAG interruption. These two types of interruptions are associated with successive maternal CTG repeat contractions and low somatic instability in DM1. Strikingly, these results suggest that a single A modification in CTG repeat expansions is sufficient to alter the CTG repeat dynamics in DM1 patients by mechanism(s) yet to be discovered.

ACKNOWLEDGMENTS

The authors would like to thank DM1 patients, Banque Genethonculture cellulaire and colleagues at *Imagine* Institute for helpful discussion and comments. We specially thank Thierry Larmonier, Eliane Gardais, Safaa Saker-Delye, Pauline Arnaud, Chasserieau Raphaële, Pascal Cintas, Ana-maria Cobo Esteban, Marie-Carmen Cruz, Dalil Hamroun, Armelle Magot, Alexandra Nadaj-Pakleza, Anne-Catherine Aube-Nathier, and Andoni Urtizberea for their active involvement in this project.

DISCLOSURE STATEMENT

The authors declare no conflict of interest.

ORCID

Stéphanie Tomé (D) http://orcid.org/0000-0002-0135-9256

REFERENCES

- Anvret, M., Ahlberg, G., Grandell, U., Hedberg, B., Johnson, K., & Edstrom, L. (1993). Larger expansions of the CTG repeat in muscle compared to lymphocytes from patients with myotonic dystrophy. *Human Molecular Genetics*, 2(9), 1397–1400.
- Ashizawa, T., Anvret, M., Baiget, M., Barcelo, J. M., Brunner, H., Cobo, A. M. ... Harley, H. (1994). Characteristics of intergenerational contractions of the CTG repeat in myotonic dystrophy. *American Journal of Human Genetics*, 54(3), 414–423.
- Ashizawa, T., Dubel, J. R., & Harati, Y. (1993). Somatic instability of CTG repeat in myotonic dystrophy. *Neurology*, 43(12), 2674–2678.
- Ashizawa, T., Dunne, P. W., Ward, P. A., Seltzer, W. K., & Richards, C. S. (1994). Effects of the sex of myotonic dystrophy patients on the unstable triplet repeat in their affected offspring. *Neurology*, 44(1), 120–122.
- Ashizawa, T., & Harper, P. S. (2006). Myotonic dystrophies: An overview. In R. D. Wells & T. Ashizawa (Eds.), Genetic instabilities and neurological disorders(2nd ed, pp. 21–36). San Diego: Elsevier. Academic Press.
- Barbe, L., Lanni, S., Lopez-Castel, A., Franck, S., Spits, C., Keymolen, K., ... Pearson, C. E. (2017). CpG methylation, a parent-of-origin effect for maternal-biased transmission of congenital myotonic dystrophy. *American Journal of Human Genetics*, 100(3), 488–505. https://doi.org/ 10.1016/j.ajhg.2017.01.033
- Bennett, R. L., French, K. S., Resta, R. G., & Doyle, D. L. (2008). Standardized human pedigree nomenclature: Update and assessment of the recommendations of the National Society of Genetic Counselors. *Journal of Genetic Counseling*, 17(5), 424–433. https://doi.org/10.1007/ s10897-008-9169-9
- Botta, A., Rossi, G., Marcaurelio, M., Fontana, L., D'Apice, M. R., Brancati, F., ... Novelli, G. (2017). Identification and characterization of 5' CCG inter-

ruptions in complex DMPK expanded alleles. *European Journal of Human Genetics*, 25(2), 257–261. https://doi.org/10.1038/ejhg.2016.148

- Braida, C., Stefanatos, R. K., Adam, B., Mahajan, N., Smeets, H. J., Niel, F., ... Monckton, D. G. (2010). Variant CCG and GGC repeats within the CTG expansion dramatically modify mutational dynamics and likely contribute toward unusual symptoms in some myotonic dystrophy type 1 patients. *Human Molecular Genetics*, 19(8), 1399–1412. doi: Ddq015 [pii] 10.1093/hmg/ddq015
- Brook, J. D., McCurrach, M. E., Harley, H. G., Buckler, A. J., Church, D., Aburatani, H. Housman, D.E. (1992). Molecular basis of myotonic dystrophy: Expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell*, 69(2), 385.
- Brunner, H. G., Bruggenwirth, H. T., Nillesen, W., Jansen, G., Hamel, B. C., Hoppe, R. L. Smeets, H.J.M. (1993). Influence of sex of the transmitting parent as well as of parental allele size on the CTG expansion in myotonic dystrophy (DM). *American Journal of Human Genetics*, 53(5), 1016–1023.
- Choudhry, S., Mukerji, M., Srivastava, A. K., Jain, S., & Brahmachari, S. K. (2001). CAG repeat instability at SCA2 locus: Anchoring CAA interruptions and linked single nucleotide polymorphisms. *Human Molecular Genetics*, 10(21), 2437–2446.
- Chung, M. Y., Ranum, L. P., Duvick, L. A., Servadio, A., Zoghbi, H. Y., & Orr, H. T. (1993). Evidence for a mechanism predisposing to intergenerational CAG repeat instability in spinocerebellar ataxia type I. *Nature Genetics*, 5(3), 254–258.
- Cleary, J. D., & Pearson, C. E. (2005). Replication fork dynamics and dynamic mutations: The fork-shift model of repeat instability. *Trends in Genetics*, 21(5), 272–280.
- De Antonio, M., Dogan, C., Hamroun, D., Mati, M., Zerrouki, S., Eymard, B., ... Bassez, G. (2016). Unravelling the myotonic dystrophy type 1 clinical spectrum: A systematic registry-based study with implications for disease classification. *Revue Neurologique*, 172(10), 572–580. https://doi.org/10.1016/j.neurol.2016.08.003
- DeJesus-Hernandez, M., Mackenzie, I. R., Boeve, B. F., Boxer, A. L., Baker, M., Rutherford, N. J., ... Rademakers, R. (2011). Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron*, 72(2), 245–256. https:// doi.org/10.1016/j.neuron.2011.09.011
- Dion, V. (2014). Tissue specificity in DNA repair: Lessons from trinucleotide repeat instability. *Trends in Genetics*, 30(6), 220–229. https://doi.org/10.1016/j.tig.2014.04.005
- Dion, V., Lin, Y., Hubert, L., Jr., Waterland, R. A., & Wilson, J. H. (2008). Dnmt1 deficiency promotes CAG repeat expansion in the mouse germline. *Human Molecular Genetics*, 17(9), 1306–1317. https://doi.org/10.1093/ hmg/ddn019
- Dogan, C., Puymirat, J., & Bassez, G. (2015). [DM-SCOPE, an intermediary appraisal report and benefits of databases in neuromuscular disorders]. *Medecine Sciences, 31 Spec No 3,* 18–19. https://doi.org/10.1051/xbrk medsci/201531s305
- Dryland, P. A., Doherty, E., Love, J. M., & Love, D. R. (2013). Simple repeatprimed PCR analysis of the myotonic dystrophy type 1 gene in a clinical diagnostics environment. *Journal of Neurodegenerative Diseases*, 2013, 857564. https://doi.org/10.1155/2013/857564
- Eichler, E. E., Holden, J. J., Popovich, B. W., Reiss, A. L., Snow, K., Thibodeau, S. N., ... Nelson, D. L. (1994). Length of uninterrupted CGG repeats determines instability in the FMR1 gene. *Nature Genetics*, 8(1), 88–94.
- Gomes-Pereira, M., Bidichandani, S. I., & Monckton, D. G. (2004). Analysis of unstable triplet repeats using small-pool polymerase chain reaction. In Y. Kohwi (Ed.), *Trinucleotide repeat protocols* (Vol. 277, pp. 61–76). Totowa: Humana Press.

WILEY Human Mutation

- Harper, P. S. (2001). The genetic basis of myotonic dystrophy. *Myotonic* dystrophy (3rd ed, pp. 307–363). London; New York: WB Saunders.
- Harper, P. S., Harley, H. G., Reardon, W., & Shaw, D. J. (1992). Anticipation in myotonic dystrophy: New light on an old problem. *American Journal of Human Genetics*, 51(1), 10–16.
- Higham, C. F., & Monckton, D. G. (2013). Modelling and inference reveal nonlinear length-dependent suppression of somatic instability for small disease associated alleles in myotonic dystrophy type 1 and Huntington disease. *Journal of the Royal Society, Interface*, 10(88), 20130605. https://doi.org/10.1098/rsif.2013.0605
- Hu, Y., Hashimoto, Y., Ishii, T., Rayle, M., Soga, K., Sato, N., ... Yokota, T. (2017). Sequence configuration of spinocerebellar ataxia type 8 repeat expansions in a Japanese cohort of 797 ataxia subjects. *Journal of the Neurological Sciences*, 382, 87–90. https://doi.org/ 10.1016/j.jns.2017.08.3256
- Huguet, A., Medja, F., Nicole, A., Vignaud, A., Guiraud-Dogan, C., Ferry, A., ... Gourdon, G. (2012). Molecular, physiological, and motor performance defects in DMSXL mice carrying >1,000 CTG repeats from the human DM1 locus. *Plos Genetics*, 8(11), e1003043. https://doi.org/ 10.1371/journal.pgen.1003043. PGENETICS-D-12-00540 [pii]
- Jarem, D. A., Huckaby, L. V., & Delaney, S. (2010). AGG interruptions in (CGG)(n) DNA repeat tracts modulate the structure and thermodynamics of non-B conformations in vitro. *Biochemistry*, 49(32), 6826–6837. https://doi.org/10.1021/bi1007782
- Jones, L., Houlden, H., & Tabrizi, S. J. (2017). DNA repair in the trinucleotide repeat disorders. *Lancet Neurology*, 16(1), 88–96. https://doi.org/ 10.1016/S1474-4422(16)30350-7
- Lavedan, C., Hofmann-Radvanyi, H., Shelbourne, P., Rabes, J.-P., Duros, C., Savoy, D., ... Junien, C. (1993). Myotonic dystrophy: Size- and sexdependent dynamics of CTG meiotic instability, and somatic mosaicism. *American Journal of Human Genetics*, 52, 875–883.
- Leeflang, E. P., & Arnheim, N. (1995). A novel repeat structure at the myotonic dystrophy locus in a 37 repeat allele with unexpectedly high stability. *Human Molecular Genetics*, 4(1), 135–136.
- Libby, R. T., Hagerman, K. A., Pineda, V. V., Lau, R., Cho, D. H., Baccam, S. L., ... 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), e1000257. https://doi.org/ 10.1371/journal.pgen.1000257
- Mahadevan, M., Tsilfidis, C., Sabourin, L., Shutler, G., Amemiya, C., Jansen, G. Korneluk, R.G. (1992). Myotonic dystrophy mutation: An unstable CTG repeat in the 3' untranslated region of the gene. *Science*, 255(5049), 1253–1255.
- Mangiarini, L., Sathasivam, K., Mahal, A., Mott, R., Seller, M., & Bates, G. P. (1997). Instability of highly expanded CAG repeats in mice transgenic for the Huntington's disease mutation. *Nature Genetics*, 15(2), 197–200.
- Martorell, L., Johnson, K., Boucher, C. A., & Baiget, M. (1997). Somatic instability of the myotonic dystrophy (CTG)n repeat during human fetal development. *Human Molecular Genetics*, 6(6), 877–880.
- Martorell, L., Martinez, J. M., Carey, N., Johnson, K., & Baiget, M. (1995). Comparison of CTG repeat length expansion and clinical progression of myotonic dystrophy over a five year period. *Journal of Medical Genetics*, 32(8), 593–596.
- Martorell, L., Monckton, D. G., Gamez, J., Johnson, K. J., Gich, I., de Munain, A. L., & Baiget, M. (1998). Progression of somatic CTG repeat length heterogeneity in the blood cells of myotonic dystrophy patients. *Human Molecular Genetics*, 7(2), 307–312.
- Martorell, L., Monckton, D. G., Sanchez, A., Lopez De Munain, A., & Baiget, M. (2001). Frequency and stability of the myotonic dystrophy type 1 premutation. *Neurology*, *56*(3), 328–335.

- Maurer, D. J., O'Callaghan, B. L., & Livingston, D. M. (1998). Mapping the polarity of changes that occur in interrupted CAG repeat tracts in yeast. *Molecular and Cellular Biology*, 18(8), 4597–4604.
- Michel, L., Huguet-Lachon, A., & Gourdon, G. (2015). Sense and antisense DMPK RNA foci accumulate in DM1 tissues during development. *Plos* One, 10(9), e0137620. https://doi.org/10.1371/journal.pone.0137620
- Monckton, D. G., Wong, L. J., Ashizawa, T., & Caskey, C. T. (1995). Somatic mosaicism, germline expansions, germline reversions and intergenerational reductions in myotonic dystrophy males: Small pool PCR analyses. *Human Molecular Genetics*, 4(1), 1–8.
- Montermini, L., Andermann, E., Labuda, M., Richter, A., Pandolfo, M., Cavalcanti, F., ... Cocozza, S. (1997). The Friedreich ataxia GAA triplet repeat: Premutation and normal alleles. *Human Molecular Genetics*, 6(8), 1261– 1266.
- Morales, F., Couto, J. M., Higham, C. F., Hogg, G., Cuenca, P., Braida, C., ... Monckton, D. G. (2012). Somatic instability of the expanded CTG triplet repeat in myotonic dystrophy type 1 is a heritable quantitative trait and modifier of disease severity. *Human Molecular Genetics*, 21(16), 3558– 3567. https://doi.org/10.1093/hmg/dds185
- Moseley, M. L., Schut, L. J., Bird, T. D., Koob, M. D., Day, J. W., & Ranum, L. P. (2000). SCA8 CTG repeat: En masse contractions in sperm and intergenerational sequence changes may play a role in reduced penetrance. *Human Molecular Genetics*, 9(14), 2125–2130.
- Mulvihill, D. J., Nichol Edamura, K., Hagerman, K. A., Pearson, C. E., & Wang, Y. H. (2005). Effect of CAT or AGG interruptions and CpG methylation on nucleosome assembly upon trinucleotide repeats on spinocerebellar ataxia, type 1 and fragile X syndrome. *Journal* of *Biological Chemistry*, 280(6), 4498–4503. doi: M413239200 [pii] https://doi.org/10.1074/jbc.M413239200
- Musova, Z., Mazanec, R., Krepelova, A., Ehler, E., Vales, J., Jaklova, R., ... Sedlacek, Z. (2009). Highly unstable sequence interruptions of the CTG repeat in the myotonic dystrophy gene. *American Journal of Medical Genetics. Part A*, 149A(7), 1365–1374. https://doi.org/ 10.1002/ajmg.a.32987
- 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), 580–588. doi: Ddq501 [pii] https://doi.org/10.1093/hmg/ddq501
- Neville, C. E., Mahadevan, M. S., Barcelo, J. M., & Korneluk, R. G. (1994). High resolution genetic analysis suggests one ancestral predisposing haplotype for the origin of the myotonic dystrophy mutation. *Human Molecular Genetics*, 3(1), 45–51.
- Nolin, S. L., Sah, S., Glicksman, A., Sherman, S. L., Allen, E., Berry-Kravis, E., ... Hadd, A. G. (2013). Fragile X AGG analysis provides new risk predictions for 45–69 repeat alleles. *American Journal of Medical Genetics. Part* A, 161A(4), 771–778. https://doi.org/10.1002/ajmg.a.35833
- Ohshima, K., Sakamoto, N., Labuda, M., Poirier, J., Moseley, M. L., Montermini, L., ... Pandolfo, M. (1999). A nonpathogenic GAAGGA repeat in the Friedreich gene: Implications for pathogenesis. *Neurology*, 53(8), 1854– 1857.
- Pearson, C. E., Eichler, E. E., Lorenzetti, D., Kramer, S. F., Zoghbi, H. Y., Nelson, D. L., & Sinden, R. R. (1998). Interruptions in the triplet repeats of SCA1 and FRAXA reduce the propensity and complexity of slipped strand DNA (S-DNA) formation. *Biochemistry*, 37(8), 2701–2708.
- Pearson, C. E., Nichol Edamura, K., & Cleary, J. D. (2005). Repeat instability: Mechanisms of dynamic mutations. *Nature Reviews Genetics*, 6(10), 729– 742.
- Pesovic, J., Peric, S., Brkusanin, M., Brajuskovic, G., Rakocevic-Stojanovic, V., & Savic-Pavicevic, D. (2017). Molecular genetic and clinical characterization of myotonic dystrophy type 1 patients carrying variant

982 WILEY Human Mutation

repeats within DMPK expansions. *Neurogenetics*, https://doi.org/ 10.1007/s10048-017-0523-7

- Polyzos, A. A., & McMurray, C. T. (2017). Close encounters: Moving along bumps, breaks, and bubbles on expanded trinucleotide tracts. DNA Repair, 56, 144–155. https://doi.org/10.1016/j.dnarep.2017.06.017
- Ramos, E. M., Martins, S., Alonso, I., Emmel, V. E., Saraiva-Pereira, M. L., Jardim, L. B., ... Silveira, I. (2010). Common origin of pure and interrupted repeat expansions in spinocerebellar ataxia type 2 (SCA2). *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics*, 153B(2), 524–531. https://doi.org/10.1002/ajmg.b.31013
- Renton, A. E., Majounie, E., Waite, A., Simon-Sanchez, J., Rollinson, S., Gibbs, J. R., ... Traynor, B. J. (2011). A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron*, 72(2), 257–268. https://doi.org/10.1016/j.neuron.2011.09.010
- Rolfsmeier, M. L., Dixon, M. J., & Lahue, R. S. (2000). Mismatch repair blocks expansions of interrupted trinucleotide repeats in yeast. *Molecular Cell*, 6(6), 1501–1507.
- Sakamoto, N., Larson, J. E., Iyer, R. R., Montermini, L., Pandolfo, M., & Wells, R. D. (2001). GGA*TCC-interrupted triplets in long GAA*TTC repeats inhibit the formation of triplex and sticky DNA structures, alleviate transcription inhibition, and reduce genetic instabilities. *Journal of Biological Chemistry*, 276(29), 27178–27187. https:// doi.org/10.1074/jbc.M101852200
- Salehi, L. B., Bonifazi, E., Stasio, E. D., Gennarelli, M., Botta, A., Vallo, L., ... Novelli, G. (2007). Risk prediction for clinical phenotype in myotonic dystrophy type 1: Data from 2,650 patients. *Genetic Testing*, 11(1), 84– 90. https://doi.org/10.1089/gte.2006.0511
- Santoro, M., Fontana, L., Masciullo, M., Bianchi, M. L., Rossi, S., Leoncini, E., ... Silvestri, G. (2015). Expansion size and presence of CCG/CTC/CGG sequence interruptions in the expanded CTG array are independently associated to hypermethylation at the DMPK locus in myotonic dystrophy type 1 (DM1). *Biochimica Et Biophysica Acta*, 1852(12), 2645–2652. https://doi.org/10.1016/j.bbadis.2015.09.007
- Santoro, M., Masciullo, M., Pietrobono, R., Conte, G., Modoni, A., Bianchi, M. L., ... Silvestri, G. (2013). Molecular, clinical, and muscle studies in myotonic dystrophy type 1 (DM1) associated with novel variant CCG expansions. *Journal of Neurology*, 260(5), 1245–1257. https:// doi.org/10.1007/s00415-012-6779-9
- Santoro, M., Masciullo, M., Silvestri, G., Novelli, G., & Botta, A. (2016). Myotonic dystrophy type 1: Role of CCG, CTC and CGG interruptions within DMPK alleles in the pathogenesis and molecular diagnosis. *Clinical Genetics*, https://doi.org/10.1111/cge.12954
- Seznec, H., Lia-Baldini, A. S., Duros, C., Fouquet, C., Lacroix, C., Hofmann-Radvanyi, H., ... Gourdon, G. (2000). Transgenic mice carrying large human genomic sequences with expanded CTG repeat mimic closely the DM CTG repeat intergenerational and somatic instability. *Human Molecular Genetics*, 9(8), 1185–1194.
- Shelbourne, P., Winquist, R., Kunert, E., Davies, J., Leisti, J., Thiele, H., ... Johnson, K. (1992). Unstable DNA may be responsible for the incomplete penetrance of the myotonic dystrophy phenotype. *Human Molecular Genetics*, 1, 467–473.
- Sicot, G., & Gomes-Pereira, M. (2013). RNA toxicity in human disease and animal models: From the uncovering of a new mechanism to the development of promising therapies. *Biochimica Et Biophysica Acta*, doi: S0925-4439(13)00076-8 [pii] https://doi.org/10.1016/ j.bbadis.2013.03.002

- Thornton, C. A., Johnson, K., & Moxley, III, R. T. (1994). Myotonic dystrophy patients have larger CTG expansions in skeletal Muscle than in leukocytes. *Annals of Neurology*, 35, 104–107.
- Tome, S., Manley, K., Simard, J. P., Clark, G. W., Slean, M. M., Swami, M., ... Pearson, C. E. (2013). MSH3 polymorphisms and protein levels affect CAG repeat instability in Huntington's disease mice. *Plos Genetics*, 9(2), e1003280. https://doi.org/10.1371/journal.pgen.1003280
- Tome, S., Nicole, A., Gomes-Pereira, M., & Gourdon, G. (2014). Nonradioactive detection of trinucleotide repeat size variability. *PLoS Currents*, 6, https://doi.org/10.1371/currents.md.ad50113b899fa13 52ce70c087eead706
- Tome, S., Panigrahi, G. B., Lopez Castel, A., Foiry, L., Melton, D. W., Gourdon, G., & Pearson, C. E. (2011). Maternal germline-specific effect of DNA ligase I on CTG/CAG instability. *Human Molecular Genetics*, 20(11), 2131– 2143. https://doi.org/10.1093/hmg/ddr099
- Usdin, K., House, N. C., & Freudenreich, C. H. (2015). Repeat instability during DNA repair: Insights from model systems. *Critical Reviews* in Biochemistry and Molecular Biology, 50(2), 142–167. https://doi. org/10.3109/10409238.2014.999192
- Volle, C. B., & Delaney, S. (2013). AGG/CCT interruptions affect nucleosome formation and positioning of healthy-length CGG/CCG triplet repeats. BMC Biochemistry [Electronic Resource], 14, 33. https://doi.org/ 10.1186/1471-2091-14-33
- Warner, J. P., Barron, L. H., Goudie, D., Kelly, K., Dow, D., Fitzpatrick, D. R., & Brock, D. J. (1996). A general method for the detection of large CAG repeat expansions by fluorescent PCR. *Journal of Medical Genetics*, 33(12), 1022–1026.
- Wong, L. J., Ashizawa, T., Monckton, D. G., Caskey, C. T., & Richards, C. S. (1995). Somatic heterogeneity of the CTG repeat in myotonic dystrophy is age and size dependent. *American Journal of Human Genetics*, 56(1), 114–122.
- Yrigollen, C. M., Martorell, L., Durbin-Johnson, B., Naudo, M., Genoves, J., Murgia, A., ... Tassone, F. (2014). AGG interruptions and maternal age affect FMR1 CGG repeat allele stability during transmission. *Journal of Neurodevelopmental Disorders*, 6(1), 24. https://doi.org/ 10.1186/1866-1955-6-24
- Zatz, M., Passos Bueno, M. R., Cerqueira, A., Marie, S. K., Vainzof, M., & Pavanello, R. C. M. (1995). Analysis of the CTG repeat in skeletal muscle of young and adult myotonic dystrophy patients: When does the expansion occur? *Human Molecular Genetics*, 4(3), 401– 406.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Tomé S, Dandelot E, Dogan C, et al. Unusual association of a unique CAG interruption in 5' of DM1 CTG repeats with intergenerational contractions and low somatic mosaicism. *Human Mutation*. 2018;39:970–982. https://doi.org/10.1002/humu.23531

SUPPLEMENTARY MATERIAL

S1. MATERIALS AND METHODS

S1.1 3' Triplet prime PCR (TP-PCR)

To analyze the purity of the CTG repeat tract at the 3' end CTG repeat array, TP-PCR was performed on both strands of the CTG repeat (Warner et al. 1996). 3' of CTG repeat was amplified using the primer downstream of the CTG repeat Somy4R-FAM (5'-FAM-CGGGTTTGGCAAAAGCAAATTTCCCGAand (5'-3') (Musova al. 2009), P3R P4CTG et TACGCATCCCAGTTTGAGACGTGCTGCTGCTGCTGCTGCT-3') primers (Warner et al. 1996). 20-100ng of DNA from blood and trophoblast were amplified in 25 μ l reaction using 0.4 μ M of FAM and P3R primers and 0.04 µM of P4CTG primer, 1× Custom master mix (Thermo Scientific) and 0.06U Thermoperfect *Taq* polymerase (Integro BV). The following cycling conditions were used: 5 min at 94°C; 60s at 94°C, 1min30 s at 68°C and 2 min at 72°C (30 cycles) and 10 min at 72°C (1 cycle). The amplified product (1-3 µl) was analyzed using 3500XL genetic analyzer (Applied Biosystems) and Gene Mapper software. GeneScan 600 LIZ dye size standard v2.0 is used in this experiment (ThermoFisher scientific).

S1.2 Interruptions analyses by *AluI* and *AciI* digestions

5 ng of DNA from blood or trophoblast were amplified in 25µl reaction using 0.4µM 101 (5'-CTTCCCAGGCCTGCAGTTTGCCCATC-3') and 102 (5'-GAACGGGGCTCGAAGGGTCCTTGTAGC-3') primers, 1× Custom master mix (Thermo Scientific) and 0.04U Thermoperfect *Taq* polymerase (Integro BV). The following cycling conditions were used: 5min at 96°C; 45s at 96°C, 30s at 68°C and 3min at 72°C (30 cycles); 1min at 68°C and 10min at 72°C (1 cycle). PCR products were purified using NucleoSpin Gel and PCR Clean-Up according to manufacturer recommendation (Macherey Nagels). 200ng of purified PCR products were digested either by AluI (NEB) recognizing AGCT site or by AciI (Thermoscientific) recognizing CCGC site. AluI digestion is combined with PCR amplification of the DMPK alleles using 101 and 102 primers and purification of expanded alleles containing CTG repeat expansion. Acil digestion is combined with PCR amplification of the DMPK alleles using 101 and 102 primers (Seznec et al. 2000). PCR products were incubated with each enzyme at 37° for 3h-16h. Digested PCR products were resolved by running 2.5% agarose gel electrophopheresis.

Supplementary figures

Sample number	Age	Expanded allele lengh (nCTG)	5' Interruption	3' Interruption
A2	49	150	Yes	No
1471	49	150	No	No
3912	45	115	No	No
5601	51	100	No	No
16007	51	150	No	No
A3	32	130	Yes	No
1201	32	130	No	No
6092	33	110	No	No
16273	35	105	No	No
15969	36	150	No	No
B2	33	310	Yes	No
2811	29	220	No	No
3692	33	210	No	No
5289	37	220	No	No
6496	26	280	No	No
4883	30	220	No	No

TABLE S1. Molecular data in A2, A3, B2 individuals and patients from classical DM1 families.



FIGURE S1. 3'TP-PCR at the *DMPK* CTG repeat locus in families A and B. The y-axis represents the intensity of fluorescence (arbitrary unit) whereas the x-axis represents the size in bp of each PCR product. Each fragment obtained by TP-PCR is represented by one peak.



FIGURE S2. Identification of CAG interruption by *AluI* enzymatic digestion in members from family A. (-) corresponds to undigested PCR product and (+) corresponds to digested PCR product. The sizes of the molecular weight standard 100bp are indicated on the left. The undigested product in (+) can be explained by somatic mosaicism or/and polymerase slippage during the CTG repeat amplification.



FIGURE S3. Identification of CCG interruption by *AciI* enzymatic digestion in members from family B. (-) corresponds to undigested PCR product and (+) corresponds to digested PCR product. Normal and expanded alleles are indicated in bracket. The sizes of the molecular weight standard 100bp and 250bp are indicated on the figure.

REFRENCES

Musova, Z., Mazanec, R., Krepelova, A., Ehler, E., Vales, J., Jaklova, R., . . . Sedlacek, Z. (2009). Highly unstable sequence interruptions of the CTG repeat in the myotonic dystrophy gene. *Am J Med Genet A, 149A*(7), 1365-1374. doi: 10.1002/ajmg.a.32987

Seznec, H., Lia-Baldini, A. S., Duros, C., Fouquet, C., Lacroix, C., Hofmann-Radvanyi, H., ... Gourdon, G. (2000). Transgenic mice carrying large human genomic sequences with expanded CTG repeat mimic closely the DM CTG repeat intergenerational and somatic instability. *Hum Mol Genet*, *9*(8), 1185-1194.

Warner, J. P., Barron, L. H., Goudie, D., Kelly, K., Dow, D., Fitzpatrick, D. R., & Brock, D. J. (1996). A general method for the detection of large CAG repeat expansions by fluorescent PCR. *J Med Genet*, *33*(12), 1022-1026.

I.5. DONNEES NON PUBLIEES

Les données qui seront évoquées ci-dessous ont été obtenues sur des culots de cellules lymphoblastoïdes de patients, immortalisées par le virus d'Epstein Barr (EBV). Les cellules nous ont été fournies par Myobank-AFM (service de banque nationale de stockage de tissus humains). Cependant, le faible nombre d'échantillons finalement disponibles ne permet que de fournir des tendances et non pas à des données significatives. Cependant, ces travaux préliminaires semblent préciser des éléments publiés dans l'article précédemment présenté.

I.5.1 Quantification de l'expression sens et antisens de DMPK

La transcription de *DMPK*, et plus particulièrement la transcription bidirectionnelle, est un facteur clef de l'instabilité des triplets CTG. C'est pourquoi nous avons cherché à quantifier les transcrits sens et antisens de *DMPK* dans des cellules issues de patients DM1 contrôles et de la famille A.

I.5.1.1. MATERIELS ET METHODES

L'extraction d'ARN ainsi que la rétro-transcription et qPCR ont été réalisées comme décrites précédemment (Huguet *et al.* 2012, Tome *et al.* 2018). Brièvement, l'extraction est réalisée par broyage mécanique (Retsch MM400) suivit d'un traitement avec le kit TRIzol Plus RNA Purification (Ambion, ref. 12183-018 Invitrogen) suivant les recommandations du fabriquant. La rétrotranscription a été réalisée sur 125 à 2000ng d'ARN avec le kit SuperScript III (Life Technologies). Les primers utilisés sont résumés dans le tableau DNP_Tableau 1. La qPCR a été réalisée avec le kit Power SYBR (Applied Biosystems) dans le thermocycler B7300 real-time PCR system (Applied Biosystems). L'intégralité du protocole a été reproduit 3 fois pour chaque échantillon.

Primers Rétrotranscription	Amorces (5'-3')		
LK- DMPK sens	CGACTGGAGCACGAGGACACTGACTTGCTCAGCAGTGTCA		
LK-DMPK antisens	CGACTGGAGCACGAGGACACTGACTTTCTTTCGGCCAGGCTGAGGC		
18S	CGGGTTGGTTTTGATCTG		
Primers Q-PCR	Amorces (5'-3')		
DMPK sens/LK	5'- CGACTGGAGCACGAGGACACTGA- 3' 5'- GGAGAGGGACGTGTTG - 3'		
DMPK antisens/LK	5'- GGAGCACGAGGACACTGA - 3' 5'- CCTTCGAGCCCCGTTCGC - 3'		
18S	5'- CGGGTTGGTTTTGATCTG - 3' 5'- CAGTGAAACTGCGAATGG - 3'		

DNP_Tableau 1: Primers utilisés pour la quantification de l'expression de DMPK

I.5.1.1. RESULTATS ET DISCUSSION

Les résultats présentés dans la figure ci-après une tendance à la diminution de l'expression de DMPK totale dans le cas de répétitions CTG interrompue (membres de la famille A) comparée aux individus contrôles (d'un facteur allant de x0,5 à x2). Cependant, la variabilité inter-individuelle couplée au faible nombre d'échantillons disponibles ne permet pas de conclure définitivement sur la diminution présumée de l'expression. Les individus les plus proches en âges et en nombre de répétitions comparables sont les individus III.1 pour la famille A, avec le contrôle DM1 12779. Si l'on compare ces 2 individus, on remarquera que la transcription antisens est réduite de 1,6 fois, et la transcription sens 1,4 fois. Ce qui laisserait supposer que la transcription antisens serait légèrement plus réduite que la transcription sens dans le cas d'une interruption CAG. Il n'est cependant pas à exclure que les contrôles DM1 ou les patientes de la famille A puissent présenter d'autres polymorphismes qui pourraient expliquer la tendance à la diminution de l'expression de DMPK. Enfin, il a été démontré que le niveau de transcription n'est pas corrélé à l'amplitude de l'instabilité somatique (Lin et al. 2009). Cependant il n'est pas à exclure qu'une diminution éventuelle de la transcription de DMPK, couplée à d'autres facteurs comme les interruptions, puisse être une des pièces d'un mécanisme par lequel les interruptions peuvent influencer l'instabilité des triplets CTG. La constitution de lignées iso-géniques avec une même taille de répétitions CTG pure et interrompues pourrait être un bon compromis pour analyser l'impact direct des interruptions sur l'instabilité des triplets CTG ainsi que d'autres facteurs d'instabilité comme l'expression de DMPK.



DMPK sens/185 Moyenne des rapports 2 Famille A Témoins DM1 1,8 1,6 DMPK sens / 185 (UA) 1,4 1.2 1 Т 0,8 0,6 0,4 0.2 0 111.1 IV.1 IV.2 12779 13997 32 ans 27 ans 21 ans 7 ans 49 ans 250 CTG 140 CTG 125 CTG 130 CTG 170 CTG

DNP_Figure 1 : Analyse du niveau d'expression de DMPK sens et antisens dans des cellules EBV de la famille A et de contrôles DM1 à répétitions pures.

I.5.2. ANALYSE DE L'EXPRESSION DE LIG1

Il a été démontré dans un modèle murin DM1 développé au laboratoire, que l'invalidation de l'expression du gène *Lig1* codant pour l'ADN LIGASE 1 inhibait les expansions de triplets CTG en faveur des contractions uniquement lors de transmissions maternelles (Tomé *et al.* 2011). De plus, ce gène est localisé sur le chromosome 19 à proximité du gène *DMPK*. Une analyse de ségrégation réalisée au laboratoire avant mon arrivée a montré que ce gène co-ségrège avec le gène *DMPK* dans la famille A. Le gène *LIG1* étant un bon candidat capable de moduler l'instabilité des triplets CTG, j'ai donc procédé à un aperçu préliminaire de l'expression protéique de ce gène dans les lignées lymphoblastoïdes des patientes de la famille A et dans des contrôles DM1.

I.5.2.1. MATERIEL ET METHODES

15µg de protéines ont été dénaturées 5 minutes à 95°C dans du Laemmli (0,125M Tris-HCl, 4% SDS, 10% Glycérol, pH 6,8) avec 5% de béta-mercaptoéthanol. Les protéines ont été déposées sur un gel SDS-PAGE d'acrylamide 8% en présence de 10µL de marqueur de taille protéique de 250 à 10kDa (PageRuler Plus Destained Protein Ladder ; ThermoScientific). Après migration (tampon de migration : 25mM Tris, 250mM Glycine, 0,1% SDS), les protéines ont été transférées sur une membrane de transfert PVDF (Perkin Elmer) (tampon de transfer : 25mM Tris, 192 mM Glycine, 0,01% SDS, 20% v/v Ethanol) pendant 2 heures à 350 mA, 100V à 4°C. Les membranes ont été bloquée 1 heure à température ambiante dans du Blotto 5% TBST, et incubées à 4°C toute la nuit avec les anticorps primaires. L'anticorps primaire permettant de révéler l'ADN LIGASE 1 est mouse GTX 18691 (Gene tex 1 :5000), celui permettant de révéler la tubuline est mouse Anti b-Tubulin T4026 (Sigma 1 :5000). Après lavages successifs de 5 minutes, 10 minutes et 15 minutes dans du TBST 1X, les membranes ont été incubées 1 heure avec l'anticorps secondaire (Sheep anti mouse-HRP Jackson Immuno Research, 1 :5000 ou Goat anti-rabbit Santa Cruz SC 2004 1 :5000) à température ambiante. Les membranes ont ensuite été lavées successivement 5 minutes, 10 minutes et 15 minutes dans du TBST 1X. Le complexe anticorps primaire-anticorps secondaire lié à la pexoxydase a été visualisé avec du ECL Plus western Blotting (Perkin Elmer). L'acquisition a été réalisée avec l'équipement Fusion ® à exposition non saturante. Le protocole a été réalisé 4 fois dans son intégralité pour chaque échantillon.

I.5.2.2. RESULTATS ET DISCUSSION

La figure DNP_Figure 2 présente les résultats de western blot pour l'expression de l'ADN LIGASE 1 dans des cellules EBV de patientes de la famille atypique porteuse d'une interruption CAG, et de témoins DM1 (la membrane présentée est représentative des répliques de l'expérience). Aucune différence significative n'est observée entre les patientes de la famille A et les témoins DM1.



LIG1/TUB Moyenne des rapports

DNP_Figure 2 : Analyse de l'expression de l'ADN LIGASE1 par western blot.

Les graphes représentent la quantification de l'expression de l'ADN LIGASE1 rapportée à l'expression de la Tubuline. La membrane présentée est un exemple représentatif des 4 réplicatats experimentaux réalisés. UA : unité arbitraire

А

l'instar de l'analyse des niveaux de transcription sens et antisens de *DMPK*, cette expérience préliminaire ne permet pas de conclure définitivement sur une potentielle différence de production de *LIG1*. Davantage d'échantillons est nécessaire. Par ailleurs, l'idéal pour ce genre d'étude serait d'avoir des cellules non immortalisées afin de pouvoir détecter des valeurs n'étant pas influencée par l'immortalisation qui peut déréguler l'expression de certains gènes. Il serait alors intéressant de comparer l'expression d'autres gènes candidats tels que *XRCC1*, *ERCC1 ou ERCC2* par exemple afin de tester si l'interruption CAG est le seul facteur modulateur de l'instabilité des triplets CTG chez cette famille atypique.

II. OPTIMISATION DU PROTOCOLE DE SMALL-POOL-PCR

I.1. ÉLEMENTS DE CONTEXTE

Un nombre croissant de maladies neuromusculaires et neurodégénératives humaines sont identifiées comme la conséquence d'expansion anormales de triplets répétés instables. Dans la population générale, ces répétitions restent stables dans les tissus et lors des transmissions intergénérationnelles. Chez les patients, les répétitions anormalement longues sont instables et généralement biaisées en faveur des expansions dans les cellules germinales et somatiques (voir (Pearson et al. 2005) pour revue). Dans certaines maladies associées au TNR, il est observé un important phénomène d'anticipation, comme dans le cas de la DM1 : les tailles de répétitions les plus grandes sont associées aux symptômes les plus sévères et à l'apparition la plus précoce (Harper et al. 1992). Dans les cellules somatiques, l'instabilité des TNRs est tissu-spécifique et s'accentue avec l'âge des patients en même temps que progressent les symptômes (Illarioshkin et al. 1994, Morales et al. 2012). Par conséquent, l'analyse de l'instabilité des TNRs est essentielle pour une meilleure compréhension de ces maladies et permettra d'ouvrir la voie à de nouvelles approches thérapeutiques, permettant d'éviter les expansions de TNRs ou d'induire des contractions jusqu'à revenir à une longueur de répétition non pathogène. Par ailleurs, une analyse précise de l'instabilité TNRs est nécessaire pour caractériser les nouveaux modèles et outils expérimentaux impactant l'instabilité.

La SP-PCR est une méthode puissante pour analyser avec précision l'instabilité des TNRs (Monckton *et al.* 1995, Gomes-Pereira *et al.* 2004, Tomé *et al.* 2014). Les dilutions en série impliquées dans cette méthode résolvent le « smear » hétérogène révélé lors de PCR standards. Cependant, les protocoles actuels prennent du temps (au moins 3 jours incompressibles) et comportent de nombreuses étapes pouvant affecter la qualité de l'expérience, telles que la qualité du transfert et l'hybridation. De plus, les protocoles actuels nécessitent de nombreux réactifs (matériel radioactif, anticorps...) et restent coûteux. Pour éliminer ces inconvénients et faciliter la démocratisation de cette méthode très sensible, j'ai développé une méthode plus simple et plus rapide : la Flash-Small-Pool PCR (FSP-PCR).

I.2. RESUME DES PRINCIPAUX RESULTATS

Brièvement, la FSP-PCR repose sur l'amplification par PCR d'un faible nombre de molécules d'ADN, avec une des deux amorces marquées avec un fluorochrome répondant aux rayons proches infra-rouge (IRD700). Le produit de PCR est ensuite résolu suite à une migration sur gel d'agarose dénaturant au NaOH.

Les protocoles de SP-PCR ne proposaient jusqu'ici que sur des méthodologies réalisables en minimum 3 jours incompressibles : le protocole décrit dans le papier suivant est réalisable dans son intégralité en 24h seulement. Les étapes de transfert et d'hybridation sont éliminées, ce qui évite de nombreux biais expérimentaux. De plus, nous avons démontré que ce protocole était tout aussi sensible que ceux qui l'ont précédé pour des tailles de répétitions allant jusqu'à 1200 CTG. Enfin, nous n'avons pas eu besoin d'optimiser l'étape de PCR à cause du marquage en fluorescence : ce protocole est donc facilement adaptable aux PCR utilisées en routine dans les laboratoires. Une des limitations de cette méthode est le biais de PCR pouvant survenir en cas de très longues répétitions. En effet, cette méthode n'optimise pas la réaction de PCR : l'usage d'amorces marquées en proche-infrarouge ne permet donc pas une meilleure amplification des répétitions. Il est donc possible qu'en cas de très longues répétitions CTG soit très probablement à recommander.

I.3. CONCLUSIONS GENERALE ET PERSPECTIVES

Pour conclure, nous fournissons ici un protocole de Small Pool PCR plus rapide et tout aussi sensible que ses prédécesseurs pour des tailles de répétitions allant au moins jusqu'à 1200 répétitions de CTG. Ce protocole réduit les coût d'expérimentation et ne comporte pas de risques biologiques tels que la radioactivité. Cependant, ce protocole est assujetti à l'efficacité de la PCR et au risque de biais de PCR en faveur de l'allèle normal en cas de très grande taille de répétitions. La prochaine étape pour pouvoir utiliser plus généralement ce protocole réside donc dans l'optimisation des conditions de PCR, comme ce qui a déjà pu être développé dans d'autres laboratoires (Chowdhury *et al.* 2016, Aeschbach *et al.* 2017)

II.4. CONTRIBUTION

J'ai co-conçu et réalisé toutes les expériences ainsi que rédigé l'article présentés ci-après.

Reports The flash-small-pool PCR: how to transform blotting and numerous hybridization steps into a simple denatured PCR

Elodie Dandelot^{1,2} & Geneviève Gourdon^{1,2}

¹Laboratory CTGDM, Inserm UMR1163, 75015 Paris, France; ²Institut Imagine, Université Paris Descartes-Sorbonne Paris Cité, 75015 Paris, France

BioTechniques 64:262-265 (June 2018) 10.2144/btn-2018-0035 Keywords: DM1 • IRDye • microsatellites instability • PCR • trinucleotide repeats

Numerous human diseases are associated with abnormal expansion of unstable trinucleotide repeats (TNRs). TNR instability mechanisms are complex, and remain only partially understood. Small-pool-PCR (SP-PCR) is the reference method to assess TNR instability. SP-PCR amplifies a low number of DNA molecules and is followed by Southern blot. However, SP-PCR remains expensive and time consuming. Here, we describe an optimized SP-PCR that can be done in a day, which reduces cost and experimental biases: the flash-small-pool PCR (FSP-PCR). This method consists of a fluorescent PCR on a few DNA molecules, followed by an alkaline gel electrophoresis revealed with a near infra-red detector system. With reduced experimental steps, cost, and time consumption, microsatellite analysis will become more accessible due to FSP-PCR.

Introduction

An increasing number of human neuromuscular and neurodegenerative diseases are due to unstable trinucleotide repeats (TNRs), including Huntington's disease (with CAG repeats) [1], myotonic dystrophy type 1 (with CTG repeats) [2-4], Fragile X Syndrome (FXS, with CGG repeats) [5] or Friedreich's ataxia (with GAA repeats) [6]. In the general population, repeats remain stable in tissues and through parental transmissions. In patients, the abnormally expanded repeats are unstable and generally biased towards expansions in both germline and somatic cells. In some TNR-associated diseases, such as myotonic dystrophy, larger repeats are associated with more severe symptoms and with an earlier age of onset (anticipation phenomenon) [7]. In somatic cells, TNR instability is age-dependent, tissue-specific and may be associated with symptom progression [8,9]. Therefore, evaluation of TNR instability is critical for prognosis and to better understand these diseases.

Triplets instability appears as a complex result of different mechanisms [7,10]. Development of different models such as mouse, bacterial, yeast, or cell models has given insight about the processes underlying TNR instability but, to date, the dynamic of repeats changes remains only partially understood [10–13]. Further analysis on TNR instability will provide clues for new therapeutic approaches to avoid repeat expansion or to induce TNR contraction down to a non-pathogenic length. Furthermore, accurate TNR instability analysis is necessary to characterize new models and new experimental tools [14–16].

SP-PCR is a powerful method to accurately analyze TNR instability [17–20]. The serial dilutions involved in this method resolve the heterogeneous smear detected by standard PCR experiments. However, current protocols are time consuming, with many steps that can affect the experiment quality such as blotting quality, and hybridization. Moreover, current protocols require many reagents and remain expensive. To eliminate this disadvantage, and to facilitate the spread of this high-sensitive method, we have developed an easier and faster method: the flash-small-pool PCR (FSP-PCR) to detect microsatellite repeatcontaining PCR products. This protocol comes without biohazard, and does not require a blotting step or hybridization. Moreover, this protocol is faster, feasible in 24 h only, while current procedures take at least 3 days.

In short, this protocol consists of a classical small-pool-PCR performed with one of the primers labeled with near infrared dye (IRD), loaded on a denaturing agarose gel, and analyzed with fluorescent analyzer, here the Odyssey[®] CLx Infrared Imaging System (LI-COR Biosciences – GmbH, NE, USA). In this study, we focused on CTG repeat expansion in DM1 patients.

Material & methods

This protocol has been adapted from previously described procedures [16–18]. The primers used are listed in Table 1. See

METHOD SUMMARY

To analyze more quickly and easily triplet nucleotides repeats, we optimized the commonly used small-pool PCR protocol using a near infrared-labeled primer and alkaline gel electrophoresis. Our protocol gives robust results, and avoids the many experimental reagents and biases of the classic small-pool PCR.

Vol. 64 No. 6 2018

Table 1. List of primers used in flash-small-pool PCR.

Primer name	Primer sequence		
Forward primer DMC	5'-AACGGGGCTCGAAGGGTCCT-3'		
Reverse labeled primer DMBR	5'-IRD700-CGTGGAGGATGGAACACGGGAC-3'		

supplementary material for a more detailed protocol.

Restriction endonuclease digestion

All DNA quantifications were done with Qubit[®] Fluorometer (ThermoFisher Scientific, MA, USA) following manufacturer recommendations. To facilitate DNA accessibility, genomic DNA was digested with 50 U of Hind III endonuclease (Fermentas, MA, USA).

FSP-PCR

Genomic digested DNA was serially diluted (1.5 ng.µ^{L-1}, 300 pg.µ^{L-1}, 60 pg.µ^{L-1}, 12 pg. µ^{L-1}) in deionized milli Rho water with 0.1 M of carrier oligonucleotide (Table 1). PCR mix was prepared as follows (final volume = 7 µl per well): 0.2 µM of unlabeled-forward primer, 0.2 µM of IRD-Reverse primer (Metabion, Planegg, Bavaria, Germany), 1X PCR buffer (ThermoFisher Scientific, ref. SM-0005; MA, USA), 5 U.µ^{L-1} of Thermo Perfect Taq (Integro, Zaandam, The



Figure 1. FSP-PCR using blood DNA from patients carrying 150 CTG and non-DM1 controls. (A) The FSP-PCR was performed with successive dilutions of a DM1 patient blood DNA (1.5 ng,µl⁻¹, 300 pg,µl⁻¹, 60 pg,µl⁻¹ and 12 pg,µl⁻¹). For each dilution, the seven lanes represent seven PCR replicates. PCR products corresponding to normal and mutant alleles are indicated in brackets. Gel agarose percentage: 1.5%. (B) FSP-PCR using blood DNA from three DM1 patients. For each patient, seven PCR replicates were performed using the 300 pg,µl⁻¹ dilution. Gel agarose percentage: 1.5%. (C) FSP-PCR using blood DNA from three non-DM1 control individuals. For each patient, seven PCR replicates were performed using the 300 pg,µl⁻¹ dilution. Gel agarose percentage: 3%. (D) Left panel: SP-PCR (seven replicates using the DNA dilution 300 pg,µl¹) revealed with a (CAG)₇ probe labeled with digoxygenin after Southern blot. Right panel: FSP-PCR (seven replicates using the DNA dilution 300 pg,µl¹) revealed directly with Odyssey[®] CLx Infrared Imaging. L1: 50–1500 bp DNA sizing standard (LI-COR Biosciences – GmbH). L2: Thermo Scientific GeneRuler 100 bp DNA Ladder (ThermoFisher Scientific). L3: DNA Molecular Weight Marker III, DIG-labeled (Roche). DM-1: Myotonic dystrophy type 1; FSP-PCR: Flash-small-pool PCR.

Vol. 64 | No. 6 | 2018

Netherlands), Deionized milli Rho water up to 6.5 μ l. Each PCR reaction was done with 0.5 μ l of each serial DNA dilution. PCR program was launched in a Thermal Thermocycler (Applied Biosystems, CA, USA): one 5-min hold at 96°C, followed by 28 cycles: 96°C (45 s), 68°C (45 s), 70°C (3 min), and a chase at 68°C (1 min) and 70°C (10 min). No PCR optimization was needed because of infra-red labeling (IRD).

Denaturing agarose gel electrophoresis of PCR products

A 20 × 40 cm 2.5% agarose gel [19] (50 nM of NaCl and 4 mM of EDTA) was pre-soaked in a cold room for at least 1 h in running buffer (NaOH 30 mM, EDTA 2 mM, volume adjusted with milli Rho water). 2 µl of the PCR reactions were used in each well. The ladder used was the 50–1500 bp DNA sizing standard (LI-COR Biosciences – GmbH, USA). The gel electrophoresis was set up as follows 30 min at 140 V, then 2 V/cm (80 V for a 20 × 40 cm gel) overnight. Gels were directly observed with the Odyssey[®] CLx Infrared Imaging System and analyzed with Image Studio[™] software (LI-COR Biosciences – GmbH).

Human DNA samples

DNA from DM1 patients were obtained from the Genethon DNA and Cell Bank in collaboration with the DM scope registry. Consent was obtained for the use of DNA samples in this study.

Results & discussion

We provide here an optimized and robust protocol to assess microsatellite instability. While the current protocols with radioactivity or digoxygenin labeling require expensive reagents and days of experimental procedures, this new protocol provides results in a much simpler way (the experimental steps are resumed to one PCR and one agarose gel migration only), and within a decreased procedure time. This protocol has been tested using serial dilution of blood DNA from a DM1



Figure 2. CTG repeat length determination using IRD-labeled leading primer and alkaline gel electrophoresis. PCR with DMC/DMBR primers on 2 ng of DNA extracted from human cerebellum samples with (A) 500 CTG repeats (1645 bp expected) or (B) 900 CTG (2845 bp expected), and (C) mice tail with 1200 CTG (3745 bp expected). L3: Thermo Scientific GeneRuler 1 kb DNA Ladder (ThermoFisher Scientific). The ladder lane was revealed with ethidium bromide while the rest of the gel with the PCR products was revealed with Odyssey[®] CLx Infrared Imaging.

patient carrying 150 CTG repeats. Figure 1A shows clearly the CTG repeat somatic mosaicism in the different dilutions as observed in classical small-pool PCR [16]. FSP-PCR was repeated with blood DNA from three different DM1 patients carrying between 150 and 200 CTG repeats. PCR amplification using 0.5 µl of the 300 pg. µl⁻¹ dilution shows again clearly the CTG repeat mosaicisms resulting from somatic instability (Figure 1B and Supplementary Figure S1). As expected, no somatic mosaicism was detected in non-DM1 individuals (Figure 1C and Supplementary Figure S1). We can note different repeat sizes for normal alleles between individuals, reflecting DMPK normal allele polymorphisms [21]. In order to verify the efficiency of FSP-PCR, we compared the CTG repeat expansion mosaicism using FSP-PCR and classical SP-PCR (Figure 1D) [18]. The magnitude of the CTG repeat instability was comparable

using both methods showing normal sensitivity with FSP-PCR. We also used various DNA samples carrying different CTG repeat expansions including DNA extracted from DM1 cerebellum samples (with 500 and 900 CTG repeats) and tail DNA from the DMSXL mouse model carrying 1200 CTG repeats. Using 2 ng of DNA, we were able to detect up to 1200 CTG repeats (Figure 2). The results obtained demonstrate that the protocol is useful for detecting a wide range of CTG repeat lengths both in human (blood and tissues) and in mice. Since no optimization was necessary in the original PCR program used with classical unlabeled primers, we assume that this protocol can be easily adapted to any usual PCR protocol [17-20].

Recently, it has been reported that ethidium bromide modifies the mobility of CAG•CTG alternative DNA structures generated by PCR during electrophoresis [22]. Using a denaturing gel and fluorescent primers, this

protocol overcomes the possible migration bias resulting from secondary DNA structures formed during PCR. Moreover, we used here IRD700 labeling, as it is empirically known to provide sharper detection than the IRD800 labeling [23]. To our knowledge, there are no IRD-labeled ladders with bands higher than 1500 bp. This problem can be easily overcome by using a usual standard and performing a TBE 1X/ethidium bromide bath after analysis with the Odyssey CLx Infrared Imaging System: fluorescent and UV pictures can be overlaid in computational post-treatment (Figure 2). Another solution can be to create your own labeled ladder as already described elsewhere [24].

In conclusion, the FSP-PCR protocol using fluorescent PCR, an alkaline gel electrophoresis, and a near Infra-red detector system is a highly sensitive, easy, reliable, and faster method for assessing microsatellite instability.

Acknowledgements

The authors thank S.Tome and the DM-Scope registry (especially Dr. G. Bassez and C. Dogan) for their help in obtaining DNA samples.

Author contributions

E.D. designed and performed the experiments and prepared the manuscript. G.G. supervised the study and corrected the manuscript.

Financial & competing interests disclosure

This study was supported by grants from AFM-Téléthon (France, AFM-Telethon grant n° 19757 to G. G.), INSERM (France), Université Paris Descartes (France) as well as a PhD fellowship from Ministère Français de la Recherche et Technologie (to E.D.). The Imagine Institute received a state subsidy managed by the National Research Agency under the "Investments for the Future" program (ANR-10-IAHU-01). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Ethical conduct of research

The authors state that they have obtained appropriate institutional review board approval or have followed the principles outlined in the Declaration of Helsinki for all human or animal experimental investigations. In addition, for investigations involving human subjects, informed consent has been obtained from the participants involved.

Open access

This work is licensed under the Creative Commons Attribution 4.0 License. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: www.future-science. com/doi/suppl/10.2144/btn-2018-0035

References

Papers of special note have been highlighted as: • of interest; •• of considerable interest

- The Huntington's Disease Collaborative Research Group. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* 72, 971–983 (1993).
- Aslanidis C, Jansen G, Amemiya C et al. Cloning of the essential myotonic dystrophy region and mapping of the putative defect. *Nature* 355, 548–551 (1992).
- Brook JD, McCurrach ME, Harley HG et al. Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. Cell 68, 799–808 (1992).
- Fu YH, Pizzuti A, Fenwick RG Jr, King J, Rajnarayan S. An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science* 255, 1256–1258 (1992).
- Ashley CT, Sutcliff JS, Kunst CB *et al.* Human and murine FMR-1: alternate splicing and translational initiation downstream of the CGG-repeat. *Nat. Gene.* 4, 244–251 (1993).
- Campuzano V, Montermini L, Moltò MD et al. Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. Science 271, 1423–1427 (1996).
- Pearson CE, Nichol Edamura K, Cleary JD. Repeat instability: mechanisms of dynamic mutations. *Nat. Rev. Genet.* 6(10), 729–742 (2005).
- Morales F, Couto JM, Higham CF et al. Somatic instability of the expanded CTG triplet repeat in myotonic dystrophy type 1 is a heritable quantitative trait and modifier of disease severity. *Human Mol. Gen.* 21(16), 3558–3567 (2012).
- 9. Illarioshkin SN, Igarashi S, Onodera O et al. Trinucleotide repeat length and rate of progression

of Huntington's disease. *Neurology* 36(4), 630–635 (1994).

- López castel A, Cleary JD, Pearson CE. Repeat instability as the basis for human diseases and as a potential target for therapy. *Nat. Rev. Mol. Cell Biol.* 11(3), 165–170 (2010).
- Usdin K, House NCM, Freudenreich CH. Repeat instability during DNA repair: Insights from model systems. *Crit. Rev. Biochem. Mol. Biol.* 50(2), 142–167 (2015).
- Gomes-Pereira M, Cooper TA, Gourdon G. Myotonic dystrophy mouse models: towards rational therapy development. *Trends Mol. Med.* 17(9), 10.1016 (2011).
- Brouwer JR, Willemsen R, Oostra BA. Microsatellite repeat instability and neurological disease. *BioEssays* 31(1), 71–83 (2009).
- Polyzos A, McMurray C. Close encounters: Moving along bumps, breaks, and bubbles on expanded trinucleotide tracts. *DNA Repair* 56, 144–155 (2017).
- Cinesi C, Aeschbach L, Yang B, Dion V. Contracting CAG/CTG repeats using the CRISPR-Cas9 nickase. *Nat. Commun.* 7, 13272 (2016).
- Monckton DG, Wong LJ, Ashizawa T, Caskey CT. Somatic mosaicism, germline expansions, germline reversions and intergenerational reductions in myotonic dystrophy males: small pool PCR analyses. *Hum. Mol. Gene.* 4(1), 1–8 (1995).
 First description of the Small-Pool PCR method.
- Gomes-Pereira M, Bidichandani SI, Monckton DG. Analysis of unstable triplet repeats using small-pool polymerase chain reaction. In: *Trinucleotide Repeat Protocols. Methods in Molecular Biology*[™]. Kohwi Y (Ed.). Humana Press, New York City, NY, USA (2004).
 • Full description of the Small-Pool PCR method.
- Tomé S, Nicole A, Gomes-Pereira M, Gourdon G. Non-radioactive detection of trinucleotide

repeat size variability. *PLoS Currents* 6 March 2014 Edition 1 (2014).

- Drouin R, Gao S, Holmquist GP. Agarose gel electrophoresis for DNA damage analysis. In: *Technologies for Detection of DNA Damage and Mutations.* Pfeifer GP (Ed.). Springer, Boston, MA, USA (1996).
- Aeschbach L, Dion V. Minimizing carry-over PCR contamination in expanded CAG/CTG repeat instability applications. *Sci. Rep.* 7(1), 18026 (2017).
- Zerylnick C, Torroni A, Sherman SL, Warren ST. Normal variation at the myotonic dystrophy locus in global human populations. *Am. J. Hum. Genet.* 56, 123–130 (1995).
- Gomes-Pereira M, Monckton DG. Ethidium bromide modifies the agarose electrophoretic mobility of CAG•CTG alternative DNA structures generated by PCR. *Front. Cell. Neurosci.* 11, 153 (2017).
- Ying B-W, Fourmy D, Yoshizawa S. Substitution of the use of radioactivity by fluorescence for biochemical studies of RNA. *RNA* 13(11), 2042– 2050 (2007).
- Gupta S, Charakana C, Sreelakshmi Y, Sharma R. Fluorescent dye labeled DNA size standards for molecular mass detection in visible/infrared range. *BMC Res Notes* 4, 12 (2011).

Received: 9 January 2018; Accepted for publication: 9 May 2018

Address correspondence to Geneviève Gourdon, Laboratory CTGDM, Inserm UMR1163, 75015 Paris, France. E-mail: genevieve.gourdon@inserm.fr

To purchase reprints of this article, contact s.cavana@future-science.com

The Flash-Small-Pool PCR Protocol

PROTOCOL FOR:

The Flash-Small-Pool PCR : How to transform blotting and numerous hybridization steps into a single PCR.

LEGEND



W REST

REAGENTS

Milli Rho Water

Ice

Sodium Chloride (Dutscher, Brumath, France)

Sucrose (Sigma-Aldrich, Saint-Louis, Missouri, USA)

Orange-G ((Sigma-Aldrich, Saint-Louis, Missouri, USA)

EDTA pH8 (Euromedex, Souffelweyersheim, France)

Sodium Hydroxide (Carlo Erba, Val-de-Reuil, France)

Ultra-Pure agarose (Invitrogen, Carlsbad, California, USA)

Taq Polymerase (Integro, City, State, Country)

PCR Master Mix (ThermoFisher Scientific, Waltham, Massachusetts, USA, ref. SM-0005)

Restriction enzyme Hind III (Fermentas, Waltham, Massachusetts, USA)

Labeled IRD700 primer in 5' (Metabion, Planegg, Bavaria, Germany, design on demand)

Eppendorf tubes 1.5 and 0.5mL (Dutscher, Brumath, France)

PROCEDURE

1- RESTRICTION ENDONUCLEASE DIGESTION

1. Since SP-PCR is highly sensitive, determine the stock DNA concentration accurately. Here we use Qubit[®] Fluorometer technology, following manufacturer recommendations.

2. Check the DNA quality by Ultra Violet/ethidium bromide agarose gel.

3. Digest 600 ng of genomic DNA in a final volume of 50μ L with 50 units (U) of chosen enzyme. For DMC/DMBR primers, use Hind III, cutting outside the PCR fragment (final DNA concentration: 12 ng. μ L⁻¹), 3 h at 37 °C. Stop the reaction 15 min at 65°C.

*

Always keep DNA solution homogenous by vortexing.



Digestions can be stored at -20 °C for one month.

2- FLASH-SMALL-POOL PCR



To ensure that serial dilutions are homogeneous, avoid small pipetting volumes

(<10µL).

4. Serially dilute digested genomic DNA with deionized water containing 0,1uM of carrier unlabeled oligonucleotide (DMC for instance). Usually start with 1.5 ng. μ L⁻¹, 300 pg. μ L⁻¹, 60 pg. μ L⁻¹ and 12 pg. μ L⁻¹.

5. Prepare PCR mix (Final Volume= 7 μ L per well): 0,2 uM of unlabeled-Forward primer, 0,2 uM of near Infra-Red-Dye labeled (IRD) Reverse primer, 1X PCR Master Mix, Thermo Perfect Taq (5 U. μ L⁻¹) Deionized bi-distilled water up to 7 μ L. Amplify 0,5 μ L of each of the serial DNA dilution in each PCR reaction.

*

To avoid contamination and serial pipetting, we recommend to prepare a master

mix and then to distribute it in wells. Prepare replicates for each dilution to appreciate mosaicism accurately (7 wells per dilution is a good starting point).

6. Launch PCR reaction: 5 min at 96 °C to denature samples. Followed by cycle reactions through 28 round of 96 °C (45 s), annealing temperature (45 s), 70°C (3 min), and a chase of annealing temperature (1 min) and 70 °C (10min).

*

We did not have to change the PCR program by using IRD labeling. Any PCR

protocols may be applicable without further optimization.

PCR can be stored at -20 °C for at least month without any degradation.

DNA dilutions at 1.5 ng. μ L ⁻¹ and 300 pg. μ L ⁻¹ can be stored at -20°C for one

week.

3- DENATURING AGAROSE GEL ELECTROPHORESIS

7. Prepare 20x40 cm neutral agarose gel: 50 mM of NaCl and 4 mM of EDTA, for 100-200 CTG repeats, 3% of agarose gives good resolution, Adjust the volume with milli Rho water up to 400 mL. Melt the mix with microwave.

8. Ensure the cooling by using magnetic agitator during 15 min.

9. Pour the mix in a gel-forming tray. Let it set for at least 1 hour.

To ensure gel percentage, check your preparation weight before and after

agarose melting and add water to replace the water evaporated during boiling

^{*}

As NaOH stops the agarose polymerization, be sure to wait enough for gel

polymerization before putting it in the denaturing buffer.

10. Prepare the Denaturing Running buffer (NaOH 30 mM, EDTA 2 mM) adjust volume with milli Rho water.

MilliQ water should be avoid as its pH is too acid and its resistivity is much more

higher: its use results in an important gel warming.

11. Let the gel soak in the Denaturing Running buffer in a cold room (4 °C) for at least 1 hour.

Gel can also be soaked overnight.

12. Put ice all around the migration apparatus to avoid gel warming,

Be sure to surround correctly the running apparatus with ice: the more the

better. Otherwise the buffer and gel will heat a lot. This mays result in both gel deformation and blured/vanished signal.

13. Mix 3 μ L of PCR replicates with 2 μ L of Loading Buffer and 5 μ L of millirho water (Final Volume =10 μ L).

14. Load the gel. Here we use the 50-1500 bp DNA Sizing Standard from Licor –Ghm.

-Normally, the orange G turns reddish because of the pH. It is crucial to use

orange dye, as blue dyes give background signal during revelation.

- An empty well can be used to put a mix of bromophenol/cyanol blue to check the run speed. Depending on the quantity of blues put, it may turns yellowish.

15. Apply an initial power of 140 V for 30 minutes to avoid sample diffusion.

16. Turn the voltage to 90 V overnight.

*

*

*

- Milli-ampere indicator should be at this step near 200 mA-250 mA.

-16 hours run is a good starting point for length as 100-200 CTG repeats

-Higher voltage can be launch for a faster migration. But constant supervision is needed to change ice and buffer to avoid the gel warming.

4- PCR PRODUCTS DETECTION

17. Check the gel warm: ice may have melt, but temperature should remains less than 10 $^\circ\text{C}.$

If using blue dye to check migration, in case of 3 % agarose gel launched 16h at

90 V, the cyanol blue should be 3c m away from the wells, the orange G may not be visible anymore.

18. Take the gel out of the tank and put it in the revelation machine of your choice. We used Licor Odyssey[®] CLx Infrared Imaging System, but this protocol might be compatible with other machine that detects 700 and 800 nm signal.

19. Cut your gel if necessary.

20. Set the focus at 2.5 mm, and determine the detection area (in « Acquire » ribbon, the « Drawing new » button).

21. Set resolution to 160 uM.

22. Do a preview to select more accurately the zone to scan to save time (about 5min).

23. Acquire definitive picture with "Start" button (about 40 min for the whole gel acquisition).

*

*

-Black and white pictures are easier to analyze than keeping the red filter on.

- Avoid bubbles between the plate and the gel which can result is a shadow in the final picture.

24. If needed the gel can be re-put in migration after acquisition.

-If the gel needs further migration, put fresh and cold Denaturing Running

buffer.

*
TABLES/FIGURES

Table 1. List of primers used in Flash-Small-Pool-PCR

Primers names	Primers sequences
Forward primer DMC	5'-AACGGGGCTCGAAGGGTCCT-3'
Reverse labeled primer DMBR C	5'-IRD700- CGTGGAGGATGGAACACGGGAC-3'

RECIPES

Water 396,5 mL - Agarose 10 g 3% NaCl 1,168 g 50 mM EDTA 0,5M 3,5 mL 4 mM *Use Milli Rho water 50 mM					
Agarose10 g3%NaCl1,168 g50 mMEDTA 0,5M3,5 mL4 mM*Use Milli Rho water					
NaCl 1,168 g 50 mM EDTA 0,5M 3,5 mL 4 mM *Use Milli Rho water <u>Denaturing Running Buffer</u> (5 L)					
EDTA 0,5M 3,5 mL 4 mM *Use Milli Rho water <u>Denaturing Running Buffer</u> (5 L)					
*Use Milli Rho water <u>Denaturing Running Buffer</u> (5 L)					
Denaturing Running Buffer (5 L)	*Use Milli Rho water				
Denaturing Running Buffer (5 L)					
Denaturing Running Buffer (5 L)					
Water 4980 mL -					
NaOH 5,97 g 30 m	Μ				
EDTA 0,5 M 20 mL 2 mM					
*Use Milli Rho water					

Loading Buffer (20 mL)

Water 19,4 mL -

Orange G	60 mg	0,3% w/v (weight/volume)
Sucrose	13 g	65% (w/v)
EDTA 0,5M	400 µL	10 mM
Tris/HCl	200 µL	10 mM

*Use Milli Rho water

*Tris/HCl pH should be between 7,5-8.

TROUBLESHOOTING

GEL DEFORMATION AFTER RUNNING

Not enough running buffer : be sure you add enough buffer to cover well the gel.

GEL HEATING

1- Voltage maybe too high. Depending on you gel thickness, you may adapt your voltage so the mA indicate at the starting point between 200-250 mA.

2- Not enough ice surrounding the migration apparatus. Put more ice, the more the best!

BLURRY SIGNAL

1- Voltage maybe too high and the signal became smeary. Depending on the gel thickness, adapt the voltage so the mA indicates at the starting point between 200-250mA.

2- Depending on the gel thickness, you may adapt the focus in the acquiring machine (more or less half of the total gel thickness).

3- Inefficient DNA denaturation. Be sure to let the gel soak in the denaturing running buffer at least 1 hour in the cold room prior to migration.

WEAK SIGNAL

1- Weak PCR efficiency/Not enough PCR product load. Try to load more PCR product volume in each well.

2- Labeled primer is damaged/too old. Be sure to protect from the light the IRDlabeled primer and to store it at -20 °C for months, -80 °C for years. Aliquot your primer to avoid freeze/thawing cycles.

EQUIPMENT

- Licor Odyssey® CLx Infrared Imaging System (LI-COR Biosciences - GmbH, Lincoln, Nebraska, USA)

- Thermal Thermocycler (Applied Biosystems, Foster City, California, USA)
- 40cm Agarose gel migration apparatus (Apelex, Evry, France)
- Generator able to deliver upon 300mA (Apelex, Evry, France)
- Qubit® Fluorometer (ThermoFisher Scientific, Waltham, Massachusetts, USA)

Supplementary figure 1.



III. ROLE(S) INTERRUPTIONS DANS L'INSTABILITE DES TRIPLETS CTG

III.1. ÉLEMENTS DE CONTEXTE

Plus d'une quarantaine de pathologies neurodégénératives humaines sont causées par des répétitions microsatellites instables (Paulson 2018). Parmi elles, la DM1 est la maladie neuromusculaire la plus fréquente chez l'adulte et est causées par des répétitions de CTG•CAG au-delà de 50 répétitions dans la région 3'UTR du gène *DMPK* (Theadom *et al.* 2014). Ces répétitions pathologiques ont tendance à s'accumuler au cours des générations jusqu'à plus de 1000 répétitions (instabilité intergénérationnelle), et au cours de la vie des patients (instabilité somatique) en même temps que s'aggravent les symptômes. Par ailleurs, au fil des transmissions dans les familles DM1, les symptômes apparaissent de plus en plus tôt avec une gravité croissante : c'est le phénomène d'anticipation (Harper *et al.* 1992).

Dans ce contexte, il est estimé que 3 à 5% des patients DM1 présentent des interruptions dans les répétitions CTG•CAG (Pešović *et al.* 2017, Cumming *et al.* 2018, Cumming *et al.* 2019, Overend *et al.* 2019). Les interruptions identifiées à ce jour sont uniques ou répétées dans les triplets CTG, et sont constituées de trinucléotides CTC, CAG, CGG et CCG (voir (Santoro *et al.* 2017) pour revue et (Cumming *et al.* 2018, Tome *et al.* 2018)). Les études menées chez ces familles suggèrent fortement que les interruptions limitent l'instabilité des triplets CTG•CAG, voire biaisent l'instabilité vers les contractions de triplets. Par ailleurs, les familles DM1 présentant des interruption(s) témoignent d'une progression plus lente de la maladie ou de symptômes amoindris (Santoro *et al.* 2017, Cumming *et al.* 2018, Pešović *et al.* 2018).

Ici, notre objectif est de démontrer si les interruptions sont nécessaires et suffisantes pour influencer l'instabilité des triplets répétés CTG•CAG et d'identifier par quels mécanismes. Nous avons d'abord utilisé des approches à base d'oligonucléotides et de produits de PCR pour déterminer si les interruptions impactent la structure secondaire des répétitions CTG•CAG et la fixation des protéines nucléaires sur les répétitions. Ensuite, nous avons dérivé un modèle cellulaire qui permet la transcription bidirectionnelle de la région 3 'UTR de *DMPK* mutant inséré dans des cellules humaines (Nakamori *et al.* 2010). Ce modèle permet d'analyser l'impact d'une seule interruption CAG sur (1) la dynamique mutationnelle de CTG•CAG répète, (2) le niveau de la transcription sens et antisens de *DMPK* 3'UTR mutant et (3) la formation d'R-loops.

III.2. RESUME DES PRINCIPAUX RESULTATS

III.2.1. INFLUENCE DES INTERRUPTIONS SUR LA STRUCTURE SECONDAIRE DES REPETITIONS

Les structures secondaires influencent l'accessibilité de l'ADN aux différents acteurs de son métabolisme (réplication, réparation, transcription), éléments vecteurs d'instabilité des triplets répétés (voir Introduction chapitre III). De plus les répétitions CTG sont connues pour former des structures type hairpin (ou épingle à cheveux), qui favorisent le recrutement du MMR et du BER stimulant l'instabilité des triplets répétés. J'ai voulu déterminer une simple interruption CAG pouvait influencer la structure secondaire des répétitions CTG. Pour cela, j'ai utilisé des produits de PCR contenant un même nombre de répétitions CTG pures ou interrompues (voir description sur la figure 1.A) que j'ai fait migrer sur gel d'acrylamide en conditions natives.

Dans ces conditions, j'ai observé des retards de migration pour chacune des natures de répétitions de CTG (pure et interrompues). Cela témoigne de la formation de structures secondaires dans toutes des natures de répétitions. Ainsi, **une unique interruption CAG en** 5' des répétitions n'empêche pas la formation de structures secondaires, mais aboutit à la formation de structures secondaires différentes de celles formées en cas de répétitions pures.

Afin de mieux caractériser les structures présentes sur les répétitions pures et interrompues, j'ai testé la présence de structures simple brin (par digestion avec l'enzyme MungBean), et de mismatches (par digestion avec l'enzyme Surveyor). Les produits des traitements enzymatiques ont ensuite été révélés par migration en conditions natives sur gel d'acrylamide, suivit d'un traitement au bromure d'éthidium. La plupart des structures secondaires identifiées ont répondu positivement aux 2 traitements enzymatiques, tant sur les répétitions pures qu'interrompues par un unique triplet CAG. Ce qui permet de conclure que la plupart des structures secondaires révélées sur les répétitions pures et interrompues comportent des boucles simples brin et des mismatches, ce qui est compatible avec la formation d'hairpins et/ou de petites boucles extra-brin. Par ailleurs, j'ai également identifié des structures résistantes à la dégradation par Surveyor et Mung Bean : l'ADN révélé par cette bande ne contient donc ni mismatches, ni d'ADN simple brin. Ces bandes sont les témoins de brins parfaitement hybridés. Les différences de migration entre les fragments porteurs de répétitions pures et interrompues non exposés aux enzymes peuvent s'expliquer par des hairpins et/ou des boucles de tailles différentes et/ou réparties différemment le long des répétitions CTG. Une analyse en microscopie électronique ou par traitement aux bisulfites couplé au séquençage Sanger pourraient être informatifs pour analyser cette répartition.

III.2.2. INFLUENCE DES INTERRUPTIONS SUR LA FIXATION DES PROTEINES NUCLEAIRES AUX REPETITIONS CTG

Il a été démontré que le métabolisme de l'ADN influence l'instabilité des triplets CTG, notamment par le recrutement du MMR sur les mismatches et glissements causés par les répétitions. En effet, il a été démontré que MSH2 et MSH3 (formant le complexe Mutsß) étaient primordiaux pour observer des expansions de triplets (voir Introduction chapitre III.4.2.3). Les interruptions étant associées à une réduction de l'amplitude des expansions dans la mosaïque somatique des patients DM1, voire à des contractions intergénérationnelles, j'ai donc testé si l'interruption CAG empêchait la fixation de facteurs, ou, si au contraire, favoriserait l'interaction des répétitions avec des facteurs supplémentaires en comparaison avec les répétitions de CTG pures. Pour cela, j'ai réalisé une analyse de retard sur gel (EMSA) à base d'oligonucléotides double brin, marqués avec un fluorochrome en proche-infrarouge, porteur de 12 répétitions pures ou interrompues avec 6 paires de bases flanquantes à chaque extrémité des répétitions (voir article tableau 1). Les oligonucléotides ont été incubés avec des extraits nucléaires de cellules HeLa, puis, le produit de cette incubation a été révélé par migration sur gel d'acrylamide en conditions natives.

Selon la nature des répétitions, différents profils de migrations sur gels ont été révélés, ce qui suggère une différence dans la fixation de protéines aux répétitions. Par exemple, les répétitions pures ont 1 bande retardée supplémentaire comparée aux répétitions interrompues par un CAG (voir article figure 2 bande B-b). Ainsi, il est possible que les protéines recrutées en fonction de la nature des répétitions soient différentes, ou bien que les protéines soient les mêmes mais avec une affinité plus ou moins forte ou induisant la formation de complexes différents selon la nature des répétitions. Par exemple, après fixation de MSH2, selon la nature du défaut reconnu, MSH3 ou MSH6 sera recruté pour former un complexe sur l'ADN anormal. Afin de savoir si les interruptions modifient le recrutement du MMR sur les répétitions CTG, j'ai réalisé un supershift qui consiste à rajouter un anticorps (ici dirigé contre MSH3) après l'incubation de l'ADN avec les protéines nucléaires (voir article figure 2). Un supershift suite au rajout de l'anticorps anti MSH3 a été observé pour toutes les natures de triplets. Ainsi l'interruption CAG n'empêche pas le recrutement du MMR, et plus précisément de MSH3 sur les répétitions CTG. Dans le profil de migration des répétitions pure un complexe ne répond pas au traitement avec l'anticorps anti MSH3, ce qui suggère qu'en cas d'interruption, la disponibilité de MSH3 pour l'anticorps utilisé est changée, et que cela peut-être dû à l'absence de certaine(s) protéine(s) ne sont plus recrutée sur

les répétitions CTG. Il sera intéressant de tester différents anticorps en différentes quantités et ciblant différents sites de MSH3 afin de confirmer ce résultat. En cas de démonstration d'un acteur tierce, il sera nécessaire de l'identifier afin de déterminer quelle est la voie métabolique les interruptions peuvent influencer influencent. Il y a aussi d'autres protéines candidates qui seront testées par supershift (MSH2, MSH6). L'utilisation de spectophotométrie de masse est aussi à l'étude.

III.2.3. IMPACT D'UNE SEULE INTERRUPTION CAG SUR L'INSTABILITE DES TRIPLETS CTG

Dans les familles DM1, les interruptions dans les répétitions CTG sont associées à une mosaïque somatique réduite, notamment dans le cas d'une seule interruption CAG en 5' des répétitions (voir résultats chapitre I). J'ai donc voulu tester si cette interruption à elle seule était suffisante pour diminuer l'amplitude de l'instabilité somatique. Pour cela, j'ai dérivé un modèle cellulaire qui consiste à intégrer aléatoirement par nucléotransfection couplée à l'action de la Phi-Intégrase, un plasmide contenant la région 3'UTR de *DMPK* humaine, flanqué de promoteurs permettant la transcription bidirectionnelle dans le génome de cellules humaines (ici la lignée HEK293). Nous avons ainsi inséré 123 répétitions CTG pures ou interrompues (avec une unique interruption CAG en 30^{ème} position en 5' des répétitions) dans des cellules HEK293 et analysé l'instabilité somatique par FSP-PCR sur 4 clones indépendants pour chaque nature de répétitions (soient 8 clones au total), cultivés en parallèles pendant 42 semaines.

Après 42 semaines de cultures, une mosaïque somatique présentant des expansions et des contractions de triplets CTG est détectable. Ce résultat est en adéquation avec ce qui avait déjà été publié sur le modèle dont nous nous sommes servi ici (Nakamori *et al.* 2010). Une mosaïque somatique est détectable dans chaque clone, indépendamment de la nature des répétitions. Cependant, les clones porteurs d'une interruption CAG ont des mosaïques somatiques d'amplitude réduite lorsqu'elles sont comparées aux mosaïques présentées par les clones à répétitions pures. L'ensemble de ces résultats démontrent qu'une unique interruption CAG n'élimine pas l'instabilité somatique, mais protège des plus grandes expansions de triplets. Par ailleurs, ce résultat est en adéquation avec ce qui avait été observé dans le sang de patientes d'une famille DM1 atypique porteuse d'une interruption CAG unique en 5' des répétitions (voir résultats chapitre I). Cela permet de renforcer l'hypothèse que, dans ces familles, l'interruption CAG est un élément critique dans l'instabilité, et donc de l'avancée de la DM1. Il sera intéressant de renouveler l'expérience avec d'autres natures d'interruptions afin de comparer leur effet protecteur contre les grandes expansions de triplets. De plus, le développement d'autres modèles comme des lignées isogéniques non

immortalisées avec insertion ciblée des répétitions pures et interrompures par Crispr/Cas9 permettra d'obtenir des conditions plus proches de ce qui peut se produire *in vivo*.

III.2.4. IMPACT D'UNE SEULE INTERRUPTION CAG SUR LE METABOLISME DES ARN *DMPK*

III.2.4.1 NIVEAU DE TRANSCRIPTION

Il a été démontré que la transcription de *DMPK* est nécessaire pour observer une instabilité des triplets CTG. De plus, la transcription bidirectionnelle favorise l'instabilité des triplets. Cependant, le niveau de transcription en lui-même n'influenceraient pas l'instabilité des TNRs (Lin *et al.* 2009, Nakamori *et al.* 2010, Adihe Lokanga *et al.* 2014). Par ailleurs, la production du transcrit antisens de *DMPK* (rCAG) est augmentée en cas de répétitions pathologiques (Gudde *et al.* 2017). Afin de déterminer si l'interruption CAG modifiait la transcription de *DMPK*, j'ai comparé par RT-qPCR la quantité de transcrits produits par la transcription en sens (rCUG) et antisens (rCAG) de *DMPK* mutant à répétitions pures et interrompues inséré dans le modèle de cellules décrit dans le paragraphe précédent.

Aucune différence entre les répétitions pures et interrompues n'a été détectée dans notre modèle pour la transcription sens (rCUG) de *DMPK* mutant. Par contre, **le modèle présente clairement une diminution de l'expression antisens de** *DMPK* **mutant en cas d'interruption CAG. En comparant les résultats des niveaux de transcription sens et antisens de** *DMPK* **mutant, il est estimé que la transcription antisens (rCAG) produit 2,5 à 8,5 fois moins de transcrits que la transcription sens (rCUG) en cas de répétitions interrompue dans notre modèle. Ce qui s'oppose aux observations faites chez les patients, où le niveau de transcription sens de** *DMPK* **est plus élevé que le niveau de transcription antisens (Gudde** *et al.* **2017). Cette dernière observation peut s'expliquer par les promoteurs de ce modèle (ROSA26 et CMV) qui ne correspondent pas aux promoteurs de** *DMPK* **retrouvés dans le génome humain. Afin de tester l'effet des promoteurs sur le niveau de transcription dans notre modèle, il sera nécessaire de générer un modèle inversant le sens des répétitions par rapport aux promoteurs.**

III.2.4.2. FORMATION DE R-LOOPS

La transcription de *DMPK* dans la DM1 est aussi marquée par la formation d'R-loops, c'est-à-dire de triplexes ADN•ARN pouvant déréguler la transcription et pouvant entrer en conflit avec la progression de la réplication, ou encore pouvant recruter des mécanismes de réparation reconnaissant ces structures comme des erreurs (voir Introduction, chapitre III.4.3.4.). Afin d'approfondir l'analyse de l'impact de l'interruption CAG sur la transcription de *DMPK* et plus particulièrement sur la formation de R-loops, j'ai comparé la quantité d'R-loops produites en cas de répétitions CTG pures et interrompues dans notre modèle cellulaire par

immunoprécipitation avec l'anticorps S9.6 spécifique des R-loops suivie de RT-qPCR. Ces expériences ont permis d'observer une diminution de R-loops en présence d'interruption CAG: la présence d'une interruption CAG serait donc suffisante pour déstabiliser la formation de R-loops. De plus, la quantité d'R-loops détectée est dans notre modèle corrélée à la quantité de transcrit *DMPK* rCAG ce qui suggère que ces 2 résultats sont dépendants l'un de l'autre. Sachant qu'il y a moins de transcrits rCAG dans le cas d'interruptions, il est possible que comme il y a moins de transcrits disponibles, il y a moins d'R-loops qui se forme par exemple. Afin de comprendre davantage l'implication spécifique des transcrit se trouve majoritairement dans les R-loops dans ce modèle. L'obtention de cette information sera déterminante pour comprendre davantage l'implication des interruptions et de la transcription bidirectionnelle, la formation d'R-loops et donc les mécanismes impliqués dans l'instabilité des triplets répétés.

III.3. CONCLUSIONS GENERALE ET PERSPECTIVES

Le travail présenté dans ce chapitre démontre pour la première fois l'impact direct d'une interruption CAG sur l'instabilité des triplets CTG en cellule humaines. De plus, mes travaux proposent un aperçu des mécanismes par lesquels les interruptions influencent les TNRs: en modifiant les structures secondaires et la fixation des facteurs nucléaires, ainsi qu'en diminuant la transcription antisens de *DMPK* rCAG et la formation de R-loops. Ces nouvelles données permettent de mieux cerner comment les interruptions identifiées chez les patients peuvent limiter la progression des maladies à TNRs. À présent, l'analyse plus approfondie de l'impact des interruptions sur l'instabilité des triplets CTG nécessite une étude d'autres motifs d'interruptions, et d'autres modèles cellulaires plus proches des conditions *in vivo* (non immortalisées, insertion ciblée par Crispr/Cas9) pour préciser davantage les mécanismes impliqués.

III.4. CONTRIBUTION

J'ai co-conçu et réalisé toutes les expériences ainsi que rédigé les manuscrits présentés ciaprès.

MECHANISMS OF CTG REPEAT STABILIZATION BY INTERRUPTIONS

Dandelot Elodie^{1,2}, Stéphanie Tomé^{1,2}, Gourdon Geneviève^{1,2*}

1 Institut Imagine, Université Paris Descartes Sorbonne Paris Cité, Paris, 75015, France

2 INSERM UMR 974, Sorbonne Université, Association Institut de Myologie, Centre de Recherche en Myologie, Paris, France

* Geneviève Gourdon Tel: +33 1 42 16 57 07;; Email: genevieve.gourdon@inserm.fr

ABSTRACT

Instability of microsatellite repeats has been involved in 40 human diseases. Among them, Myotonic Dystrophy Type 1 (DM1) is a neuromuscular disease characterized by strong anticipation. DM1 is caused by expanded CTG•CAG repeats in the 3'UTR of *DMPK*. DM1 CTG•CAG expansions are unstable from one generation to another and in tissues with a strong bias towards expansions. In general, there is a good correlation between abnormal repeat length and symptoms severity. Recently, studies on DM1 families strongly pathed the way to demonstrate the impact of interrupted motifs on trinucleotide repeats instability, but the mechanisms involved remain unknown. Here, we provide further evidence that secondary structures of double strand CTG•CAG repeats and nuclear proteins binding profiles are affected by a single interruption in the repeat expansions (~110 CTG•CAG). In addition, we show that even a single CAG•CTG interruption at the 5' end is sufficient to reduce somatic instability in a human cell model allowing bidirectional transcription of the repeat. This stabilization is associated with a decrease in the formation of R-loops and with a decrease of the CAG antisense transcription. Our results indicate that R-loops formation and transcription are a key factor for the mechanisms of CTG•CAG expansions stabilization by interruptions motif.

INTRODUCTION

There is over 40 human genetic neurodegenerative pathologies caused by unstable repeated microsatellites (1). Among them, Myotonic dystrophy type 1, also known as DM1 or Steinert's disease, is an autosomal dominant neuromuscular disorder, described in the early twentieth century. It's estimated prevalence is 1/20 000 worldwide, making DM1 the most common neuromuscular disease in adults (2). DM1 has highly variable phenotype as well as a very pronounced anticipation. Anticipation is defined by a worsening and an earlier appearance of symptoms over successive generations. DM1 is a multi-systemic disease associated with myotonia, progressive neuromuscular degeneration, neurological and digestive disorders, ophthalmic and cardiac abnormalities, or insulin resistance (3).

DM1 is caused by an abnormal expansion of CTG•CAG triplets in the 3' Untranslated Region (3' UTR) of the DMPK gene (Dystrophy Myotonic Protein Kinase). In the general population, the number of CTG•CAG triplets is polymorphic, between 5 and 37 repeats and remains generally stable (4). In DM1 patients, the size of the repeat is expanded over 50 CTG•CAG and the number of repeats increase through generations in most cases: up to several thousand repeats can be observed in the most severe form of the disease (intergenerational instability). Expansion size variations through generations depends on the sex of the transmitting parent and on the size of the repeat in the transmitting parents (5,6). Usually, the largest intergenerational expansions are observed in maternal transmissions, which transmit more often the most severe disease form (7-9). In DM1 families, a good correlation, although not perfect, is found between the size of the repeats and the severity of the symptoms (10,11). Several parameters such as DNA repair, replication and bidirectional transcription and formation of R-loops (triplexes, DNA • RNA) are known to promote CTG•CAG repeats instability (12-17). Moreover, DNA secondary structures (such as hairpins), repeat length and purity (pure or interrupted repeat) are also known to impact the dynamics of Trinucleotides Repeats (TNRs) (18-20).

In DM1 families, the frequency of interrupted CTG•CAG repeats had been recently estimated at 3-5%, and they can occur in 5', 3' or in the middle of the repeat (19-22). Several studies strongly suggest an impact of the interruption on CTG•CAG repeats instability in patients' cells (19,22-26). So far, single or numerous CTC, CAG, CGG and CCG interruptions have been described in patients, and are linked to atypical stabilization or even to contractions in tissues and/or after transmissions. We previously reported DM1 families with interruptions (single CAG or 3 scattered CGG in 5' of the repeats), associated with stabilization or contractions of the repeats over generations. In these families, the repeats were atypically transmitted over several successive maternal transmissions and without anticipation. We also observed that the somatic mosaicism was reduced in these families compared to matched controls carrying pure repeats (22).

Here, our aim is to demonstrate whether interruptions are necessary and sufficient to affect CTG•CAG repeats instability and to get clues on the mechanisms involved. We first used *in vitro* approaches based on double strand oligonucleotides to study whether interruptions have an impact on CTG•CAG secondary structures and nuclear proteins binding. Then, we took advantage of a previously described model that allows bidirectional transcription of the 3 'UTR of *DMPK* into human cells (13) to analyze the

impact of a single CAG interruption on (1) the mutational dynamic of CTG•CAG repeats, (2) the level of the *DMPK* 3'UTR sense and antisense transcripts, and (3) the R-loops formation.

MATERIAL AND METHODS

Plasmids construction

Plasmid have been fully described before (13). Plasmids were kindly offered by the Charles A. Thornton laboratory. Briefly, we amplified by PCR the DMPK 3'UTR with primers supplemented with BgIII and Sac1 enzyme restriction sites. BgllI and Sac1 had unique cutting site in dedicated sequence in LC15-BD plasmid (see primers BgIII and SacI in table XX): plasmid and PCR products were digested with BgIII and Sac1 (Thermofisher; FD0083 and FD1133). Digested PCR products and plasmid were purified on agarose gel with Nucleospin PCR Cleaning kit (Macherey Nagel; 740609.50) according to manufacturer recommendation. PCR product was ligated on LC15-BD plasmid with DNA ligase T4 (Promega; M1801) overnight at 4°C according to manufacturer recommendation. Ligated plasmid was integrated in E.coli DH5-alpha (Thermofisher; 18265017) according to manufacturer recommendation. Clones were selected with 123 CTG repeats pure or interrupted with a single CAG triplet in 30th popistion in 5' of the repeat by sequencing (see CTG•CAG repeat sequencing paragraph) and plasmids purified with the Nucleospin Plasmid Miniprep kit (Macherey Nagel; 740588.10) according to the manufacturer recommendations. pLC15-BD with the human DMPK 3'-UTR (nt 2096-2841, accession number NM 004409) with different CTG•CAG repeat natures as described in Table 1. Constitutive forward and reverse transcription of the DMPK 3'UTR insert is mediated by CMV/ chicken beta-actin enhancer/promoter and ROSA promoter respectively (figure 3A). LC15-BD plasmid also contains puromycin resistance and attB element recognized by PhiC31 for integration in the HEK293 host genome. We co-nucleotransfected pLC15-BD with PhiC31o plasmid encoding the phiC31 integrase (gifts from Thornton CA. 's lab) to mediate integration of pLC15-BD at pseudo-attP sites in HEK296 cells genome.

Secondary structures on gel assay

We generated sequences by PCR on plasmids carrying pure or interrupted CTG•CAG repeats and described above. 6 pmol of plasmid DNA was amplified in a 25 µl reaction using 0.4 µM DMC and DMBR primers, in 1 × Custom master mix (Thermo Fischer Scientific; SM005) and 0.04U Thermoperfect Taq polymerase (Peak International Product b.v, Eerbeek). Cycling conditions: 5 min at 96°C; 45 sec at 96°C, 30 sec at 60°C and 3 min at 72°C (30 cycles); 1 min at 60°C and 10 min at 72°C (1 cycle). PCR products were purified with PCR cleaning Nucleospin kit (Macherey Nagel; 740609.50) according to manufacturer protocol, excepting that the final incubation with 1X TE buffer for elution was longer : 10min. 50 ng of purified PCR products were annealed in 1X TE buffer (final volume 25uL) using a thermocycler (95°C 5min then -1°C/min until 25°C). To test presence of mismatches, 400 ng of PCR products were digested with specific enzyme Surveyor (IDT; 706025) according to manufacturer recommendation. To test the presence of single strand DNA, 1ug of PCR product were digested with Mung Bean nuclease (Progmega; M4311) following manufacturer recommendation. All samples were

loaded in the same 5% native acrylamide gel with 100pb and 1kbp ladder (Thermofisher; SM0241 and SM1331). Samples runned in gel overnight at 4°C in 1X TBE Buffer.Gel, the gel was revealed with Ethidium bromide and GelDoc (Biorad) imager.

CTG repeat binding assay

To check the pure and interrupted CTG•CAG repeats binding to nuclear proteins, we used double strand 48 pb oligonucleotides (CTR, INT CAG, see Table 1) labelled with near infrared probe (Metabion). 100 pmol of forward oligonucleotides were annealed with 100 pmol of non-labelled complementary reverse oligonucleotide (OLELO_1R and OLELO_2R respectively) in 1X TE (final volume 10μ L), with thermocycler (95°C 5min then -1°C/min until 25°C). To check the annealing secondary structure, 10 pmol of annealed product were runned in 12% native acrylamid gel for 4 h at 4°C. Electrophoretic Mobility Assay (EMSA) was done with Li-Cor EMSA kit (Li-Cor; 829-07910), with 10 pmol of annealed primers mixed with 5 µg of whole nuclear protein extract from HeLa Cells (Santa Cruz; sc-2120) in the following reaction buffer: 1X Binding buffer from the kit, 2,5 mM DTT/ Tween20, 1 µg Poly (dl.dC), 0,1% NP-40, 10 mM MgCl2, final final volume of 20 µL adjusted with ultra-pure water. Reaction was incubated 40 min at room temperature in the dark. For supershift assay, 1 µg of MSH3 antibody was added after the 40 min of incubation, and samples incubated 15 min more. The whole reaction was launched in a 5% native acrylamide gel in 0.5X TBE buffer at room temperature for 1h. Gels were revealed with Odyssey Li-cor Odyssey Clx fluorescence scanner (Li-Cor)

Cell culture, nucleotransfection and selection

The cells have been mainly generated as decribed before (13). Briefly, HEK293 cells were cultured in DMEM (Gibco; 31966047) supplemented with 20% fetal bovine serum (Thermofisher; 10270106), 1% of Pennicillin and 1% streptamycin (Thermofisher;15070063). 2.105 cells were co-nucleotransfected with 80 ng of CTG-repeats containing plasmid and 800 ng of plasmid PhiC310. Random nucleotransfection was done with the kit SE Cell Line 4DX kit (Lonza; V4XC-2032) and Nucleofector 4D programm CM-130 (Lonza). For each CTG•CAG repeat nature, 4 reactions of nucleotransfections were done simultaneously and splitted in P6 wells plate (finally, 18 wells per CTG•CAG repeat nature). After 2 weeks under 50 µg.mL⁻¹ puromycin selection (Euromedex; SIH-248-50), positive wells were separated in monoclonal way by FACS. Monoclonality was checked by whole well video microscopy (Incucyte) for 1 week immediately after FACS selection. We selected at least 4 clones per CTG•CAG repeat nature and following 42 weeks culture time was under constant puromycin selection.

CTG•CAG repeat sequencing

CTG•CAG repeats nature in plasmids and cell clones were verified by sequencing. For genomic DNA on HEK293 cells, specific mutant *DMPK* transgene was amplified by PCR with 0.4 µM PLLCBD_F_ED and PLLCBD_F_ED primers (upstream and downstream of the repeats), 1× Custom master mix (ThermoScientific; SM005) and 0.04 U Thermoperfect Taq polymerase (Integro BV). The following cycling conditions were used: 5 min at 96°C; 45 s at 96°C, 30 s at 60°C and 3 min at 72°C (30 cycles); 1 min at 60°C and 10 min at 72°C (1 cycle). PCR products were purified using NucleoSpin Gel PCR

Clean-Up (Macherey Nagels;b 740609.50) following manufacturer recommendations, excepted the last elution incubation we made last 10 min. We sequenced the 40 ng of PCR products with BigDye Terminator v3.1 cycle sequencing kit (Applied Biosystems; 4337455) and ST300F or ST300R primers with the following program: CM-130. Plasmids were directly sequenced with ST300F and ST300R primers with the following program: 1 cycle of denaturation 96° (5 min) and 30 cycles of 96° (30 s), 58° (15 s), 60° (45 s). Sequence reactions were sequenced with 3500xL Genetic analyser (Applied Biosystem) and analysed with SnapGene® software (from GSL Biotech; available at snapgene.com).

CTG•CAG repeat instability analysis

Flash Small pool PCR (FSP-PCR) amplifications and PCR product electrophoretic analyses were performed using the methods previously described (27). Briefly, DNA samples from HEK293 cells were extracted by phenol-chlorophorm, quantified by Qubit Fluorometer assay (Life technology) and digested with HindIII enzyme. FSP-PCR was performed using DM-C and near-Infrared labelled premier IRD-DM-BR primers (Metabion). Up to 50 equivalent genome molecules number were denatured by heating to 96°C (5 min). PCR reactions involved 28 cycles of 96°C (45 sec), 68°C (45 sec), and 70°C (3 min) with a chase of 1 min at 68°C and 10 min at 70°C. FSP-PCR products were loaded on 2,5% agarose gel in denaturing NaOH condition at 90V for 16hr. Gel were revealed using a method based on the Odyssey Li-cor Odyssey Clx fluorescence scanner (Li-Cor).

Bidirectionnal DMPK expression quantification

RNAs were extracted from HEK293 cells as previously described (28). Briefly, cells were homogenized in Trizol (Thermo Fischer Scientific; 15596018) using beads ruptor. Total RNA was purified using PureLink RNA miniKit (Ambion, Thermo Fischer Scientific; 12183018A) according to the manufacturer's protocol. To analyze the specific mutant *DMPK* sense and antisense expression, cDNA was synthesized using Superscript III reverse transcriptase (Thermo Fischer Scientific; 18080044) and strand specific primers (Primer sense: LK-Nak RT-Sense and primer antisens: LK-Elo_AS_nak_ R) from 750 ng of RNA from HEK293 cells, according to manufacturer's protocol. RNAs were tested with retrotranscriptase (RT+) and without retrotranscriptase (RT-) to verify the absence of DNA genomic contamination. Normal HEK293 cells (without nucleotransfected mutant *DMPK*) was used as negative control for PCR specificity.

R-loops immuno-precipitation

DNA/RNA hybrids were isolated mainly as described before (29). Whole nucleic acids from cells nucleotransfected and normal HEK293 were extracted using NucleoSpin Tissue kit (Macherey-Nagel; 740952.50) according to manufacturer's protocol. Samples were sonicated with Bioruptor (30 s ON, 90 s OFF, 6 cycles) at the concentration of 20 ng. μ L⁻¹ to 400 pb fragments. Free RNA was eliminated by RNase I digestion (37°C for 1 h) followed by phenol-chloroform extraction. Samples at this point were used as input. 3 µg of sonicated samples were immunoprecipitated in Immuno-Precipitation Buffer (IP Buffer: 10 mM Na-Phosphate pH 7.0, 0.14 M NaCI, 0.05% Triton X-100) and 10 µg of S9.6 antibody

RESULTS

Interruptions in CTG•CAG repeats modulate DNA secondary structures

We used native migration on polyacrylamide gel coupled with specific enzymatic digestion to test if a single CAG interruption can influence CTG repeat secondary structures. Indeed, secondary structures are known to influence DNA metabolism such as replication, repair, and transcription, mechanisms known to be involved in trinucleotide repeats instability. CTG repeats are known to form hairpin structures, which promote the recruitment of MMR and BER stimulating the instability of repeated triplets (18,31,32). To compare secondary structures between pure and interrupted repeats, we assay reannealed PCR products (obtained with primers DMC/DMBR) in conditions allowing DNA secondary structures in native polyacrylamide gel, and treated DNA with specific enzymes recognizing mismatches (Surveyor) or simple strand and top of extra-strand loops (Mung Bean) (figure 1C). The PCR products contain CTG repeat with 34bp in 5' of the repeats and 113bp in 3': if CTG repeats would have been without flanking sequences, slippage may had occurred and provided artefactual results. Both PCR products show migration retardations, and most of the DNA retarded was digested by enzymes: only a single band resist on the treatment independently of the CTG repeat nature. Interestingly, the profiles of migration delays differ between pure and interrupted repeats, demonstrating that a single CAG interruption is able to modulate CTG repeat secondary structures. Vanishing bands after enzymatic treatment reveals that most of the pure and interrupted CTG repeats secondary structures contain mismatches and single strand, and the remaining band witness perfect double strand structures. Combination of mismatches and extra-strand loops is compatible with the hairpin formation, so a single CAG interruption may not disrupt hairpin formation.

The CAG interruption of CTG•CAG repeats affects nuclear factors binding in vitro

We used gel retardation assay (EMSA) to investigate if the CAG interruption could modify the binding of nuclear factors such as MSH2 and MSH3 (forming the MutSß complex), known to bind CTG repeats and to promote expansions (33-36). Furthermore, the interruption could create a new binding site for a factor that could impact the formation of secondary structures and/or the dynamic of the repeat. Labelled double strand oligonucleotides (12 pure CTG repeats and 12 CAG interrupted CTG repeats surrounded by 5 bp in 5' and 7bp in 3' from de DMPK sequence) were annealed in condition allowing secondary structures and incubated with Hela nuclear extracts. Both oligonucleotides showed a major band (B-a, figure 2) that was "shift" with the MSH3 antibody suggesting that this band correspond to the oligonucleotides binding MutSß (as MSH3 is stable only when complexed with MSH2 (37)) (figure 2). Two lower and fainter bands (B-b and B-c) are also observed with the pure CTG repeat but are barely detectable with the interrupted repeat. The upper band B-b disappeared in the supershift experiment with MSH3 antibody suggesting also that the band is also containing MutSß. The lower band B-c as not affected by the MSH3 antibody. It could represent the binding of MutSa (note : doit être testé prochainement avec l'anticorps MSH6). No specific band was observed with the interrupted sequence suggesting that the potential new binding site generated by the CAG interruption does not bind protein at least in Hela cells.

The single CAG interruption is sufficient to reduce CTG•CAG repeat instability in cells.

We previously showed that DM1 patients carrying about 100 CTG repeat interrupted by a single CAG present a reduced somatic mosaicism compared to matched DM1 "controls" carrying pure repeat (22). In order to test if the single nucleotide change (T replaced by A in one trinucleotide) was directly responsible for the stabilization of the trinucleotide repeat sequences, we used HEK293 cells to derived a cellular model carrying integrated plasmid with the 3'UTR region of *DMPK*, flanked by promoters allowing bidirectional transcription (figure 3A). We inserted plasmids carrying 123 pure CTG repeat or interrupted CTG repeats (with CAG interruption at the 30th position in 5 ' as observed in the DM1 family studied previously). Instability of the CTG repeat was investigated by FSP-PCR in 4 independent clones for each type of repeat (8 clones in total), grown in parallel for 42 weeks. After 42 weeks of culture, we clearly observed somatic mosaics composed of CTG triplet expansions and contractions in every clones carrying pure repeats (figure 3B). Strikingly, the extent of mosaicism was clearly reduced in clones with interrupted repeats demonstrating that a single nucleotide change in one trinucleotide can affect somatic instability and reduce the size of somatic CTG repeat length changes.

CTG·CAG sense an antisense transcription

It has been shown that transcription of trinucleotide repeats expansions is necessary to observed instability of the repeats in models (38-40). Moreover, bidirectional transcription promotes triplet instability on CTG·CAG, although the level of transcription itself doesn't influence the TNRs instability (13,39,41). In order to investigate if the CAG interruption modifies the bidirectional transcription of the *DMPK* 3'UTR in our cell model, we used specific RT-qPCR to compare the amount of transcripts produced by "sense" transcription (CTG sense) and "antisense" transcription "CAG sense" after 6 and 42 weeks of culture (Figure 4). Both sense and antisense transcripts were observed in clones carrying pure or interrupted repeats. Furthermore, We detected no difference in the levels of sense transcripts between pure or interrupted repeats (figure 4B). However, we observed a significant decrease of antisense transcripts in cells carrying interrupted repeats, although transcription levels vary between clones (figure 4A and 5B).

CAG interruption affects the formation of R-loops formation at the CTG·CAG site.

Bidirectional transcription of *DMPK* at the DM1 locus is known to produce R-loops, that promotes instability (42). Indeed, R-loops can affect transcription by promoting antisense transcription, can be conflictual with replication progression, or may recruit repair mechanisms that recognize these structures as errors (17,30,43-45). In order to further analyse the impact of the CAG interruption on the 3'UTR *DMPK* transcription and more particularly on the R-loops formation, we compared the quantity of R-loops produced in the context of pure and interrupted CTG repeats by immunoprecipitation with the R-loops-specific S9.6 antibody followed by RT-qPCR. R-loops were detected with both pure and interrupted repeats (figure 4D, 5A). However, the levels of R-loops were significantly decreased in presence of the CAG interruption. Furthermore, we observed that the R-loops amounts detected are correlated with the level of antisense transcription (figure 5), suggesting that these two results are dependent on each other.

DISCUSSION

Several studies in DM1 patients have already strongly suggested a stabilizing role of the interruptions (19,22,46). However, the direct impact of the interruptions on the CTG repeats instability and the mechanisms involved remained unknown.

Secondary structures are known to be involved in TNRs instability and CAG and CTG are both known to produces hairpins and loops structures: a melting point analysis study demonstrated that CTG and CAG repeats more frequently form multiple small hairpins than one large one (47,48). Here, we provide evidences that even a single CAG interruption in the 5' of ~100-150 CTG repeats is able to modulate TNRs repeats secondary structures. Structures detected in both pure and interrupted repeats involved single strand and mismatches which are compatible with hairpins and extra-loops formation. The differences between the pure and interrupted secondary structures more likely reside either in the shape of the structures or in their number and localization over the CTG repeat. Knowing that secondary structures are determinant for trinucleotides instability, this result provides new clue on a potential starting point of the interruption(s) stabilizing impact (47,49-51).

In this context, we demonstrated by EMSA that the CAG interruption modulates proteins recruitment on the CTG repeats. As the MMR, and more especially MSH3, is a key factor for CTG repeats instability, we tested the MSH3 recruitment on the pure and interrupted repeats by supershift. We show that MSH3 is recruited differently over pure and interrupted repeats. This result suggest that the CAG interruption may induce variation on the MMR recruitment over the repeats. On the other hand, this result can also witness the interruption's ability to stop other factor(s) recruitment which can forbid MSH3 recognition by the antibody used in this study case of pure repeat. Our results may a first evidence of a determinant interruption's mechanism stabilizing the TNRs, which would involve the MMR, key actor of CTG repeats instability.

Moreover, we demonstrated, that a single interruption is sufficient to observe a stabilization of the interruption. Indeed, we showed that in case of CAG interruption in 5' of the CTG repeats, the TNRs are protected against large repeat number variations in a human cell model that we developed. This cell model is in line with the reduced somatic mosaicism observed in DM1 patients with a single CAG interruption: a reduced mosaic amplitude when compared to DM1 controls with pure repeats. (22). So, we provide here a pertinent model to analyze the direct impact of interruptions on CTG repeats instability. It is important to note here that our model, clones present an instability with both CTG repeats expansion and contractions, contrary to the expansion bias usually saw in patients and inter-clonal variability, which is a well-known characteristic observed in cell cultures models (52-54).

In parallel with the reduced mosaicisms, and despite inter-clonal variability, we observed a significant decreased of the antisense rCAG transcript levels in the clones carrying the interrupted repeat but not of the sense rCUG transcripts suggesting an impact of the interruption only on the rCAG transcript. Moreover, the rCAG transcript (antisense) level is 2 to 8 times more important than the rCUG (sense) transcript one. This last observation is contradictory with data in literature. Indeed, it is known that antisense rCAG transcripts level is lower than the sense rCUG one (55). This may be due to the plasmid

construction as constitutive bidirectional transcription of mutant *DMPK* is led by CMV (sense rCUG transcription) and ROSA26 (antisense rCAG transcription) promoters which don't correspond to the native *DMPK* promoters. This suggestion is even more fairly plausible as it has been demonstrated that ROSA26 promoter promotes transcription stronger than CMV promoter in mice cells (56).

We show that R-loops formation is reduced in clones with interruption, suggesting that a single interruption is enough to decrease R-loops formation or to destabilize them. Interestingly, rCAG transcript levels and R-loops level detection are correlated in our model. This observation strongly suggests an interdependence of the rCAG transcript and the R-loops formation over CTG repeat nature influence (pure or interrupted). So far, two hypotheses can explain this result. First, as the level of rCAG transcript is reduced in the interrupted clones, there is less RNA that can form R-loops, so general Rloops level is reduced in interrupted clones. Secondly, it has been already described in human cells that R-loops formed in 3' of genes induce antisense transcription (57,58).So, in our cells, there may be less R-loops formed because of the interruption, so R-loops stimulate less the antisense rCAG transcription and the rCAG transcripts level is reduced in case of interruption. Previous studies in plasmid didn't detect any difference in R-loop formation between pure or interrupted repeats with in vitro transcription on plasmids (17). However, we can suppose that the different repeat nature (only 49 CAG/CTG•CAG repeats, and multiple ATG interruptions) and the plasmid in vitro transcription by polymerases T6, T7 are differences that can explain our divergent results. By the way, these differences in results may provide some clues on the interruption's impact modulators: CTG repeat length, interruption nature, or kind of RNA polymerases for instance.

Depending on the predictions used, the Phi integrase has about one hundred possible recognition sites in human genome, without evident selection bias (13,59): as each clone presented here comes from independent nucleotransfection assay, it is likely each clone may have a different integration site and it is known that genomic environment impact CTG instability (60,61). Even so, we clearly detected mosaic amplitude reduction, significant rCAG and R-loop levels decrease in case of interruption. Thus, we provide here direct evidence that the repeat nature impact on the CTG repeat instability is stronger than the insertion environment. Moreover, insertion site impact may also be reduced thanks to the 5.3kbp flancking sequences on the 5' of the CTG repeats and 1.2kpb on 3'. Still, impact of the insertion site can't be completely ignored, and may partly explain the inter-clones variability.

To conclude, our work provides new insights into the mechanisms by which interruptions can influence the instability of repeated triplets: secondary structure formation, different recruitment of nuclear proteins on CTG repeats which may include the MMR, reduction of the rCAG transcription and of the formation of R-loops in a human cell model. Altogether, this data can let us speculate on a global scheme of interruptions impact ways on CTG repeat instability. First, the CAG interruption modulates CTG repeats secondary structures. The proteins recruitment over the repeat would be modulated by the structure or by the CAG motif itself. Among the proteins recruited differently, there would be the MMR actors and/or secondary structures and/or R-loops processing., Secondary structures, proteins differently recruited and/or the CAG motif itself altogether lead to a reduced antisense rCAG transcripts level in correlation with the R-loops level decrease. R-loops may be reduced because of a reduced amount of RNA

available and/or by their destabilization by the interruption (processing, structure stability...). R-loops are known to stimulate TNRs instability, their reduction thanks to the interruption may be a leading mechanism to the interruption stabilizing properties.



Model of CAG impact on CTG repeats instability.

Our study has demonstrated that a single interruption on CTG repetitions influences the secondary structure of the repeats, the recruitment of nuclear proteins (notably MSH3), and reduces the level of transcription of the rCAG transcript in correlation with the formation of R-loops (full arrows). It is reasonable to assume (dashed arrows) that variations in secondary structures and proteins recruitment on repeats may influence DNA Transcription, Replication, and Repair over the repetitions which are known to modulate CTG instability. Moreover, modulation of the level of R-loops (decrease in case of CAG interruption) may also influence the Replication and Repair of DNA mechanisms by competition variation between the different DNA metabolism machineries.

This work has paid particular attention to the single CAG interruption but it will be interesting to reproduce this kind of approach on other types of interruptions identified in DM1 families. In particular the CCG interruptions that were more frequently identified in families of DM1 patients and which are known to form different secondary structures than hairpins: the G-quadruplex (62). A better understanding of the instability of repeated triplets is essential to better understand TNRs diseases and develop appropriate therapeutic strategies. In recent years, the increase of observations of atypical CTG instability and decrease of the symptoms with the presence of interruptions underlines the importance of interruptions in the TNR instability domain, and especially in their stabilization or contractions, phenomenon whose mechanisms are currently little known.

FUNDING

Institut National de la Santé et de la Recherche Médicale; Université ParisDescartes; AFMTelethon (16331 and 19757); National Research Agency (ANR-10-IAHU-01).

CONFLICT OF INTEREST

Authors declare no conflict of interest.

TABLE AND FIGURES LEGENDS

Table 1. Sequences of oligonucleotides used in this study.

The oligonucleotides for Gamma Actine R-loop detection were described in (Groh, Lufino et al. 2014).

Experiment	Oligonucleotide name	Oligonucleotide sequence (5' to 3')
DMPK 3'UTR amplification for	Bgl2-1	GGTGGTAGATCTTTTATTCGCGAGGGTCGGGGGGTGG
bacteria transformation	Sac1-1	GGTGGTGAGCTCGCGCCGCCCAGGAGCCG
Flash-Small-Pool PCR	DMC	AACGGGGCTCGAAGGGTCCT
	IRD- DMBR	IRD- CGTGGAGGATGGAACACGGAC
Bidirectional transcription of <i>DMPK</i> transgene	LK - Elo_AS_nak_ R	CGACTGGAGCACGAGGACACTGA - CCTCGGTATTTATTGTCTGTCCC
	Elo_AS_nak_F	GGCTTGGTGCGTTTGCGG
	LK - Nak RT-Sense	CGACTGGAGCACGAGGACACTGA - TCTAGAGAATAGGAACTTCGGAATAG
	Nak_Sense2	GCAACGTGCTGGTTATTGTG
	LK	CGACTGGAGCACGAGGACACTGA
	hPol2ra - F	AGAGAAGCTGGTGCTCCGTA
	hpol2ra - R	AGCGCAGGAAGACATCATCA
R-loops detection	CTCF 1bM2	CCTTCGAGCCCCGTTCGC
	AMH6	CGCCGCCCAGGAGCCGC
	R-loop Gamma actine F	CCGCAGTGCAGACTTCCGAG
	R-loop Gamma actine R	CGGGCGCGTCTGTAACACGG
Secondary structures on PCR product	DMC	AACGGGGCTCGAAGGGTCCT
	DMBR	CGTGGAGGATGGAACACGGAC
Secondary structures by bisulfite treatment	PLLCBD_F_ED	GGGCAACGTGCTGGTTATTG
	PLLCBD_R_ED	GCTAGGTAGGGGATCGGGAC
	PLLC- bisulf F	TGTTTTATTATTTTGGTAAAGAATT
	PLLC- bisulf R	AAATCTAAACTTAAAATCTTTTATTC
	SP6	ATTTAGGTGACACTATAG
	Τ7	TAATACGACTCACTATAGGG
	CTR OLELO_1_F	GGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGGGGGGGA
EMSA	CTR OLELO_1_R	TCCCCCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA
LINGA	INT CAG OLELO_2_F	GGAATGCTGCTG CTG CTGCTGCTGCTGCTGCTGCTGCTGGGGGGGA
	INT CAG OLELO_2_R	TCCCCCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA
Native <i>DMPK</i> Sequencing and clones screeening	ST300-F	GAACTGTCTTCGACTCCGGG
	ST300-R	GCACTTTGCGAACCAACGAT

Figure 1. Impact of the CAG interruption on mutant CTG•CAG repeats structures

(A) Scheme of the PCR products amplified with DMC/DMBR primers used in C pannel. (B) Scheme of hairpins structures Structure on the left represents hairpins in cruciform way, center scheme represents various little hairpins along the repeat, which is the most common form observed in CTG/CAG repeats. (C) Native polyacrylamide gel migration of reduplexed PCR products with pure and interrupted CTG•CAG. Lanes under Surveyor assay show mismatches detection, lanes under Mungbean assay show single strand detection. (-) without enzyme, (+) enzyme, (L1) for 100pb Ladder, (L2) for 1kpb ladder.





Figure 2. Impact of the CAG interruption on proteins binding over the CTG•CAG repeats EMSA and supershift experiment with OLELOs double strand oligonucleotides, HeLa nuclear extract proteins and MSH3 antibody (see legend on each well). B-n exhibit band of interest representing delayed complexes migration gel; (+) means "with" and (-) means "without" (see lines label); "MSH AB" means Antibody anti MSH3.



Figure 3. A Single CAG interruption reduce CTG•CAG repeat instability in human cells

(A) Scheme of the plasmid inserted in HEK293 cells. The AttB site lead to random integration in human genome, "puro" means puromycin resistance gene which permit puromycin selection in cell culture, CMV/CBA and ROSA promote sense and antisense transcription of the DMPK 3'UTR respectively. (B) Flash-Small-Pool PCR of pure and interrupted CTG•CAG repeats in HEK293 cells. For each CTG•CAG repeat nature (pure or carrying a single CAG interruption), each clone comes from an individual independent electroporation experiment. After 42 weeks cultured simultaneously, genomic DNA was extracted and about 50 molecules were amplified with IRD labelled primers in 5 replicates for each clone. "CTR" clones correspond pure CTG•CAG repeat; "INT" clones correspond to cells with a single CAG interruption; "C" correspond to the amplification of the repeats initially present in the plasmid nucleotransfected in CTR cells; "L" correspond to the ladder; "I" correspond to the amplification of the repeats initially present in the plasmid nucleotransfected in INT cells; "B" corresponds to PCR blank.



Figure 4. Impact of the CAG interruption on DMPK mRNA metabolism

RT-qPCR on cells with pure or interrupted CTG•CAG repeats. Each experiment was done twice. (A) Detection of the specific transgene sense transcription reported to the global Pol2ra transcripts. (B) Detection of the specific transgene antisense transcription reported to the global Pol2ra transcripts. (C) Evaluation of the Gamma actin R-loops fraction over total gamma actin transcripts. This data was used as a reference for the S9.6 immunoprecipitation efficiency. (D) Detection of total DMPK R-loops rationalized on immunoprecipitation efficiency. "CTR" clones correspond pure CTG•CAG repeat, "INT" clones correspond to cells with a single CAG interruption.



Figure 5. Impact of the CAG interruption on DMPK R-loops formation

(A) Detection of total DMPK R-loops rationalized on immunoprecipitation efficiency. "CTR" clones correspond pure CTG•CAG repeat, "INT" clones correspond to cells with a single CAG interruption.
 (B) Detection of the specific transgene antisense transcription reported to the global Pol2ra transcripts; (C) Detection of the specific transgene sense transcription reported to the global Pol2ra transcripts.



REFERENCES

- 1. Paulson, H. (2018), *Handbook of clinical neurology*. Elsevier, Vol. 147, pp. 105-123.
- 2. Theadom, A., Rodrigues, M., Roxburgh, R., Balalla, S., Higgins, C., Bhattacharjee, R., Jones, K., Krishnamurthi, R. and Feigin, V. (2014) Prevalence of muscular dystrophies: a systematic literature review. *Neuroepidemiology*, **43**, 259-268.
- 3. Bird, T.D. (2018) Myotonic Dystrophy Type 1.
- 4. Imbert, G., Kretz, C., Johnson, K. and Mandel, J.-L. (1993) Origin of the expansion mutation in myotonic dystrophy. *Nature Genetics*, **4**, 72-76.
- 5. Lavedan, C., Hofmann-Radvanyi, H., Shelbourne, P., Rabes, J.P., Duros, C., Savoy, D., Dehaupas, I., Luce, S., Johnson, K. and Junien, C. (1993) Myotonic dystrophy: sizeand sex-dependent dynamics of CTG meiotic instability, and somatic mosaicism. *Am J Hum Genet*, **52**, 875-883.
- 6. Ashizawa, T., Anvret, M., Baiget, M., Barcelo, J.M., Brunner, H., Cobo, A.M., Dallapiccola, B., Fenwick, R.G., Jr., Grandell, U., Harley, H. *et al.* (1994) Characteristics of intergenerational contractions of the CTG repeat in myotonic dystrophy. *Am J Hum Genet*, **54**, 414-423.
- 7. Harley, H.G., Rundle, S.A., MacMillan, J.C., Myring, J., Brook, J.D., Crow, S., Reardon, W., Fenton, I., Shaw, D.J. and Harper, P.S. (1993) Size of the unstable CTG repeat sequence in relation to phenotype and parental transmission in myotonic dystrophy. *American journal of human genetics*, **52**, 1164-1174.
- 8. Harper, P. (2009) *Myotonic dystrophy*. OUP Oxford.
- 9. Harper, P.S., Harley, H.G., Reardon, W. and Shaw, D.J. (1992) Anticipation in myotonic dystrophy: new light on an old problem. *American journal of human genetics*, **51**, 10-16.
- 10. Yum, K., Wang, E.T. and Kalsotra, A. (2017) Myotonic dystrophy: disease repeat range, penetrance, age of onset, and relationship between repeat size and phenotypes. *Current opinion in genetics & development*, **44**, 30-37.
- De Antonio, M., Dogan, C., Hamroun, D., Mati, M., Zerrouki, S., Eymard, B., Katsahian, S. and Bassez, G. (2016) Unravelling the myotonic dystrophy type 1 clinical spectrum: A systematic registry-based study with implications for disease classification. *Revue neurologique*, **172**, 572-580.
- 12. Pearson, C.E., Edamura, K.N. and Cleary, J.D. (2005) Repeat instability: mechanisms of dynamic mutations. *Nature Reviews Genetics*, **6**, 729-742.
- 13. Nakamori, M., Pearson, C.E. and Thornton, C.A.J.H.m.g. (2010) Bidirectional transcription stimulates expansion and contraction of expanded (CTG)•(CAG) repeats. *Human molecular genetics*, **20**, 580-588.
- 14. Chi, L.M. and Lam, S.L. (2005) Structural roles of CTG repeats in slippage expansion during DNA replication. *Nucleic Acids Res*, **33**, 1604-1617.
- 15. Cleary, J.D., Tome, S., Lopez Castel, A., Panigrahi, G.B., Foiry, L., Hagerman, K.A., Sroka, H., Chitayat, D., Gourdon, G. and Pearson, C.E. (2010) Tissue- and age-specific DNA replication patterns at the CTG/CAG-expanded human myotonic dystrophy type 1 locus. *Nat Struct Mol Biol*, **17**, 1079-1087.
- 16. Jones, L., Houlden, H. and Tabrizi, S.J. (2017) DNA repair in the trinucleotide repeat disorders. *The Lancet Neurology*, **16**, 88-96.
- 17. Reddy, K., Tam, M., Bowater, R.P., Barber, M., Tomlinson, M., Nichol Edamura, K., Wang, Y.H. and Pearson, C.E. (2011) Determinants of R-loop formation at convergent bidirectionally transcribed trinucleotide repeats. *Nucleic Acids Res*, **39**, 1749-1762.
- 18. Liu, G., Chen, X., Bissler, J.J., Sinden, R.R. and Leffak, M. (2010) Replicationdependent instability at (CTG)•(CAG) repeat hairpins in human cells. *Nature Chemical Biology*, **6**, 652.
- 19. Cumming, S.A., Hamilton, M.J., Robb, Y., Gregory, H., McWilliam, C., Cooper, A., Adam, B., McGhie, J., Hamilton, G., Herzyk, P. *et al.* (2018) De novo repeat interruptions are associated with reduced somatic instability and mild or absent clinical features in myotonic dystrophy type 1. *Eur J Hum Genet*, **26**, 1635-1647.

- 20. Santoro, M., Masciullo, M., Silvestri, G., Novelli, G. and Botta, A. (2017) Myotonic dystrophy type 1: role of CCG, CTC and CGG interruptions within DMPK alleles in the pathogenesis and molecular diagnosis. *Clinical Genetics*, **92**, 355-364.
- 21. Pešović, J., Perić, S., Brkušanin, M., Brajušković, G., Rakočević-Stojanović, V. and Savić-Pavićević, D. (2017) Molecular genetic and clinical characterization of myotonic dystrophy type 1 patients carrying variant repeats within DMPK expansions. *neurogenetics*, **18**, 207-218.
- 22. Tomé, S., Dandelot, E., Dogan, C., Bertrand, A., Geneviève, D., Péréon, Y., group, D.M.c.s., Simon, M., Bonnefont, J.-P., Bassez, G. *et al.* (2018) Unusual association of a unique CAG interruption in 5' of DM1 CTG repeats with intergenerational contractions and low somatic mosaicism. *Human mutation*, **39**, 970-982.
- 23. Musova, Z., Mazanec, R., Krepelova, A., Ehler, E., Vales, J., Jaklova, R., Prochazka, T., Koukal, P., Marikova, T., Kraus, J. *et al.* (2009) Highly unstable sequence interruptions of the CTG repeat in the myotonic dystrophy gene. *American Journal of Medical Genetics Part A*, **149A**, 1365-1374.
- 24. Santoro, M., Masciullo, M., Pietrobono, R., Conte, G., Modoni, A., Bianchi, M.L., Rizzo, V., Pomponi, M.G., Tasca, G., Neri, G. *et al.* (2013) Molecular, clinical, and muscle studies in myotonic dystrophy type 1 (DM1) associated with novel variant CCG expansions. *J Neurol*, **260**, 1245-1257.
- 25. Botta, A., Rossi, G., Marcaurelio, M., Fontana, L., D'Apice, M.R., Brancati, F., Massa, R., G Monckton, D., Sangiuolo, F. and Novelli, G. (2017) Identification and characterization of 5' CCG interruptions in complex DMPK expanded alleles. *Eur J Hum Genet*, **25**, 257-261.
- 26. Pešović, J., Perić, S., Brkušanin, M., Brajušković, G., Rakočević-Stojanović, V. and Savić-Pavićević, D. (2018) Repeat Interruptions Modify Age at Onset in Myotonic Dystrophy Type 1 by Stabilizing DMPK Expansions in Somatic Cells. *Front Genet*, **9**, 601-601.
- 27. Dandelot, E. and Gourdon, G. (2018) The flash-small-pool PCR: how to transform blotting and numerous hybridization steps into a simple denatured PCR. *BioTechniques*, **64**, 262-265.
- 28. Huguet, A., Medja, F., Nicole, A., Vignaud, A., Guiraud-Dogan, C., Ferry, A., Decostre, V., Hogrel, J.-Y., Metzger, F., Hoeflich, A. *et al.* (2012) Molecular, physiological, and motor performance defects in DMSXL mice carrying >1,000 CTG repeats from the human DM1 locus. *PLoS genetics*, **8**, e1003043-e1003043.
- 29. Nadel, J., Athanasiadou, R., Lemetre, C., Wijetunga, N.A., Ó Broin, P., Sato, H., Zhang, Z., Jeddeloh, J., Montagna, C., Golden, A. *et al.* (2015) RNA:DNA hybrids in the human genome have distinctive nucleotide characteristics, chromatin composition, and transcriptional relationships. *Epigenetics Chromatin*, **8**, 46-46.
- 30. Groh, M., Lufino, M.M.P., Wade-Martins, R. and Gromak, N. (2014) R-loops Associated with Triplet Repeat Expansions Promote Gene Silencing in Friedreich Ataxia and Fragile X Syndrome. *PLOS Genetics*, **10**, e1004318.
- 31. Pearson, C.E., Edamura, K.N. and Cleary, J.D.J.N.R.G. (2005) Repeat instability: mechanisms of dynamic mutations. *Nature Reviews Genetics*, **6**, 729.
- 32. Jarem, D.A., Wilson, N.R. and Delaney, S. (2009) Structure-Dependent DNA Damage and Repair in a Trinucleotide Repeat Sequence. *Biochemistry*, **48**, 6655-6663.
- 33. Foiry, L., Dong, L., Savouret, C., Hubert, L., Te Riele, H., Junien, C. and Gourdon, G. (2006) Msh3 is a limiting factor in the formation of intergenerational CTG expansions in DM1 transgenic mice. *Human genetics*, **119**, 520-526.
- 34. Guo, J., Gu, L., Leffak, M. and Li, G.M. (2016) MutSbeta promotes trinucleotide repeat expansion by recruiting DNA polymerase beta to nascent (CAG)n or (CTG)n hairpins for error-prone DNA synthesis. *Cell Res*, **26**, 775-786.
- Lai, Y., Budworth, H., Beaver, J.M., Chan, N.L.S., Zhang, Z., McMurray, C.T. and Liu,
 Y. (2016) Crosstalk between MSH2-MSH3 and polβ promotes trinucleotide repeat expansion during base excision repair. *Nature communications*, 7, 12465-12465.

- 36. van den Broek, W.J.A.A., Nelen, M.R., Wansink, D.G., Coerwinkel, M.M., te Riele, H., Groenen, P.J.T.A. and Wieringa, B. (2002) Somatic expansion behaviour of the (CTG)n repeat in myotonic dystrophy knock-in mice is differentially affected by Msh3 and Msh6 mismatch–repair proteins. *Human Molecular Genetics*, **11**, 191-198.
- 37. Acharya, S., Wilson, T., Gradia, S., Kane, M.F., Guerrette, S., Marsischky, G.T., Kolodner, R. and Fishel, R. (1996) hMSH2 forms specific mispair-binding complexes with hMSH3 and hMSH6. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 13629-13634.
- 38. Dion, V. and Wilson, J.H. (2009) Instability and chromatin structure of expanded trinucleotide repeats. *Trends in Genetics*, **25**, 288-297.
- 39. Lin, Y., Hubert Jr, L. and Wilson, J.H. (2009) Transcription destabilizes triplet repeats. *Molecular Carcinogenesis: Published in cooperation with the University of Texas MD Anderson Cancer Center*, **48**, 350-361.
- 40. Liu, G. and Leffak, M. (2012) Instability of (CTG)n•(CAG)n trinucleotide repeats and DNA synthesis. *Cell & Bioscience*, **2**, 7.
- 41. Adihe Lokanga, R., Zhao, X.-N., Entezam, A. and Usdin, K. (2014) X inactivation plays a major role in the gender bias in somatic expansion in a mouse model of the fragile X-related disorders: implications for the mechanism of repeat expansion. *Human molecular genetics*, **23**, 4985-4994.
- 42. Lin, Y., Dent, S.Y., Wilson, J.H., Wells, R.D. and Napierala, M. (2010) R loops stimulate genetic instability of CTG[.] CAG repeats. *Proceedings of the National Academy of Sciences*, **107**, 692-697.
- 43. Freudenreich, C.H. (2018) R-loops: targets for nuclease cleavage and repeat instability. *Current Genetics*, **64**, 789-794.
- 44. Groh, M. and Gromak, N. (2014) Out of balance: R-loops in human disease. *PLoS genetics*, **10**, e1004630.
- 45. Loomis, E.W., Sanz, L.A., Chédin, F. and Hagerman, P.J. (2014) Transcriptionassociated R-loop formation across the human FMR1 CGG-repeat region. *PLoS genetics*, **10**, e1004294-e1004294.
- 46. Santoro, M., Fontana, L., Masciullo, M., Bianchi, M.L.E., Rossi, S., Leoncini, E., Novelli, G., Botta, A. and Silvestri, G. (2015) Expansion size and presence of CCG/CTC/CGG sequence interruptions in the expanded CTG array are independently associated to hypermethylation at the DMPK locus in myotonic dystrophy type 1 (DM1). *Biochimica et Biophysica Acta (BBA) Molecular Basis of Disease*, **1852**, 2645-2652.
- 47. Gacy, A.M., Goellner, G., Juranić, N., Macura, S. and McMurray, C.T. (1995) Trinucleotide repeats that expand in human disease form hairpin structures in vitro. *Cell*, **81**, 533-540.
- 48. Petruska, J., Arnheim, N. and Goodman, M.F. (1996) Stability of intrastrand hairpin structures formed by the CAG/CTG class of DNA triplet repeats associated with neurological diseases. *Nucleic acids research*, **24**, 1992.
- 49. Chastain, P.D., 2nd, Eichler, E.E., Kang, S., Nelson, D.L., Levene, S.D. and Sinden, R.R. (1995) Anomalous rapid electrophoretic mobility of DNA containing triplet repeats associated with human disease genes. *Biochemistry*, **34**, 16125-16131.
- 50. Owen, B.A., Yang, Z., Lai, M., Gajec, M., Badger, J.D., 2nd, Hayes, J.J., Edelmann, W., Kucherlapati, R., Wilson, T.M. and McMurray, C.T. (2005) (CAG)(n)-hairpin DNA binds to Msh2-Msh3 and changes properties of mismatch recognition. *Nat Struct Mol Biol*, **12**, 663-670.
- 51. Gacy, A.M. and McMurray, C.T. (1998) Influence of hairpins on template reannealing at trinucleotide repeat duplexes: a model for slipped DNA. *Biochemistry*, **37**, 9426-9434.
- 52. Kang, S., Jaworski, A., Ohshima, K. and Wells, R.D. (1995) Expansion and deletion of CTG repeats from human disease genes are determined by the direction of replication in E. coli. *Nat Genet*, **10**, 213-218.
- 53. Pearson, C.E., Nichol Edamura, K. and Cleary, J.D. (2005) Repeat instability: mechanisms of dynamic mutations. *Nat Rev Genet*, **6**, 729-742.

- 54. Pelletier, R., Krasilnikova, M.M., Samadashwily, G.M., Lahue, R. and Mirkin, S.M. (2003) Replication and expansion of trinucleotide repeats in yeast. *Mol Cell Biol*, **23**, 1349-1357.
- 55. Gudde, A.E.E.G., van Heeringen, S.J., de Oude, A.I., van Kessel, I.D.G., Estabrook, J., Wang, E.T., Wieringa, B. and Wansink, D.G. (2017) Antisense transcription of the myotonic dystrophy locus yields low-abundant RNAs with and without (CAG)n repeat. *RNA Biology*, **14**, 1374-1388.
- 56. Chen, C.-m., Krohn, J., Bhattacharya, S. and Davies, B. (2011) A comparison of exogenous promoter activity at the ROSA26 locus using a ΦiC31 integrase mediated cassette exchange approach in mouse ES cells. *PloS one*, **6**, e23376-e23376.
- 57. Sollier, J. and Cimprich, K.A.J.T.i.c.b. (2015) Breaking bad: R-loops and genome integrity. *Trends in cell biology*, **25**, 514-522.
- 58. Skourti-Stathaki, K. and Proudfoot, N.J. (2014) A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. *Genes Dev*, **28**, 1384-1396.
- 59. Chalberg, T.W., Portlock, J.L., Olivares, E.C., Thyagarajan, B., Kirby, P.J., Hillman, R.T., Hoelters, J. and Calos, M.P. (2006) Integration specificity of phage phiC31 integrase in the human genome. *J Mol Biol*, **357**, 28-48.
- 60. Gourdon, G., Dessen, P., Lia, A.S., Junien, C. and Hofmann-Radvanyi, H. (1997) Intriguing association between disease associated unstable trinucleotide repeat and CpG island. *Annales de genetique*, **40**, 73-77.
- 61. Gourdon, G., Radvanyi, F., Lia, A., Duros, C., Blanche, M., Abitbol, M., Junien, C. and Hofmann-Radvanyi, H. (1997) Moderate instability of a 55 CTG repeat in transgenic mice carrying a 45 kb genomic region from an affected DM patient. *Nat Genet*, **15**, 190-192.
- 62. Zhao, X.-N. and Usdin, K. (2016) Ups and Downs: Mechanisms of Repeat Instability in the Fragile X-Related Disorders. *Genes (Basel)*, **7**, 70.

III.5. DONNEES NON DESTINEES A LA PUBLICATION

III.5.1. IMPACT D'UNE INTERRUPTION UNIQUE CAG SUR L'INSTABILITE DES TRIPLETS CTG DANS DES CELLULES HUMAINES NON IMMORTALISEES

Initialement, le modèle cellulaire décrit par Nakamori et *al.* nécessite une culture de cellules humaines non immortalisées MRC-5 (cellules de poumon de fœtus humain). Cependant, les auteurs analysaient des répétitions de triplets beaucoup plus longues (800 répétitions) que celles étudiées ici (123 répétitions) : en effet nous avons pris le parti d'analyser des tailles de répétitions au plus proches de ce qui avait pu être observé chez la famille décrite dans Tomé *et al. 2018*.

J'ai en première instance ré-utilisé le modèle de Nakamori et *al.* à l'identique, à l'exception de la longueur des répétitions CTG que nous avons choisie à 123 répétitions. La figure présentée ci-dessous correspond à une Small-Pool PCR sur environs 30 molécules d'ADN selon le protocole décrit par (Tomé *et al.* 2014) après 20 semaines de cultures



DNP_Figure 1: Small-Pool PCR sur clones MRC-5 avec des répétitions CTG pures ou interrompues.

M.XX signifie « MRC5-XX » ; « CTR » signifie Contrôles à répétitions pures ; « INT » signifie que les cellules étaient porteuses de répétitions CTG interrompues par une interruption CAG.

Les clones présentés sont un échantillon représentatif des 5 clones sélectionnés pour chacune des natures de triplets (pure ou interrompue). Après 20 semaines de cultures, une

très faible instabilité était détectable uniquement dans les clones à répétitions pures. Malgré ce résultat encourageant, le protocole ainsi constitué n'a pas pu être poursuivi. En effet, les cellules se multipliaient de moins en moins après 12 semaines de cultures, et entraient rapidement en sénescence une fois dépassées 20 semaines en cultures. C'est pourquoi nous avons par la suite choisi de poursuivre l'analyse dans des cellules immortalisées HEK293 qui avaient déjà été utilisée dans l'analyse de l'instabilité des triplets répétés (Lin *et al.* 2010, Lin *et al.* 2010, Lai *et al.* 2013)

III.5.2. DETECTION DES R-LOOPS SUR FIBROBLASTES PRIMAIRES

Au cours de l'étude présentée dans les paragraphes précédents, les quantification d'Rloops ont été réalisées sur cellules HEK293. Or, ces cellules étant immortalisées, j'ai voulu vérifier que les R-loops détectées avec notre protocole n'était pas uniquement la conséquence de l'immortalisation et/ou de l'insertion du plasmide dans le génome. Pour cela, j'ai testé le protocole dans des fibroblastes primaires de patients DM1 porteur d'environs 1000 CTG et dans des fibroblastes d'un individu normal en suivant la même stratégie que celle décrite dans le paragraphe III.2.4. Des R-loops ont été détectées dans chaque type de cellules, cependant, le niveau de R-loops détecté dans les cellules DM1 était 90 fois supérieur à celui détecté dans les cellules normales. Ce résultat démontrant la robustesse du protocole pour la détection de la dérégulation de la formation de R-loops en cas d'expansion pathologique dans des cellules de patients, il paraissait d'autant plus important d'analyser l'impact des interruptions sur le métabolisme des R-loops dans le modèle cellulaire décrit dans les paragraphes précédents.



DNP_Figure 2: Quantification de R-loops sur le gène DMPK dans des fibroblastes primaires DM1 et normaux.

III.5.3 IMPACT DES INTERRUPTIONS CCG SUR LE RECRUTEMENT DE PROTEINES NUCLEAIRES SUR LES REPETITIONS DE CTG

Le résultat de l'expérience suivante n'a pas été inclus dans l'article à ce jour faute de réplicas expérimentaux pour cette nature d'interruptions. L'analyse de l'impact de ces interruptions est toujours en cours dans le laboratoire.

Selon la nature des interruptions, il est possible que de nouveau sites de fixation de protéines apparaissent, ou, au contraire, disparaissent. J'ai ainsi testé le recrutement de protéines sur des oligonucléotides porteur de différentes interruptions CCG pour comparer l'impact des différents types d'interruptions par une expérience d'EMSA et de supershift avec l'anticorps anti-MSH3 comme décrite dans le paragraphe II.2.2.

Sur la suivante, on peut voir différents retards sur gel pour les 2 natures de répétitions interrompues : les interruptions CCG n'empêchent donc pas la fixation de protéines nucléaires. Selon la nature de l'interruption CCG, les complexes sont différents. En effet, le motif INT(CCG)₃ présente 1 complexe (B-e) supplémentaire que pour le motif INT(CCG)₂. Par ailleurs, le retard B-a est plus marqué dans les interruptions INT(CCG)₂ est plus marqué que dans la migration de INT(CCG)₃ malgré un signal global plus de plus faible intensité. Ces différences entre les profils de migration suggèrent que selon le nombre d'interruptions CGG la formation de complexes sur les répétitions CTG peux varier. Ainsi, il est probable que l'impact des interruptions sur l'instabilité des triplets n'emprunte pas tout à fait les mêmes mécanismes en fonction de leur nombre. De plus, avec l'ajout d'anticorps anti-MSH3 dans le milieu réactionnel, on observe un reliquat de l'ensemble des complexes non fixé par l'anticorps pour les répétitions avec 3 interruptions, tandis que la totalité des complexes ont été reconnus par l'anticorps en cas INT(CCG)2. Ceci peut-être dû à (1) une quantité d'oligonucléotides plus importante dans le milieu réactionnel avec INT(CCG)₃ par rapport au milieu avec INT(CCG)₂ (2) à une affinité plus forte de MSH3 pour les répétitions INT(CCG)₃ comparé aux répétitions INT(CCG)₂, ce qui mène en cas de INT(CCG)₃ à une saturation de l'anticorps, auquel cas l'ensemble des complexe n'a pas pu être fixé par l'anticorps. Cette dernière théorie pourrait être révélatrice de davantage de mésappariments de base à cause de l'élévation du nombre d'interruptions CCG, ce qui pourrait être la conséquence de la formation de structures secondaires différentes.



DNP_Figure 3:Figure 25: Impact de différents motifs d'interruptions CCG sur le recrutement de protéines sur les répétitions CTG/CAG.

EMSA et Supershift sur des oligonucléotides double brins mélangés à des extraits nucléaires HeLa en présence d'un anticorps anti-MSH3 (voir selon légende). B-n met en avant des retards sur gels représentatifs de complexes protéines-oligonucléotides ; « + » signifie « avec », « -« signifie sans protéines et/ou anticorps (voir légende). Séquences des oligonucléotides :

Il est donc nécessaire de reproduire l'expérience, et, à l'instar de ce qui a pu être présenté pour l'interruption CAG, il sera informatif de tester des anticorps anti-MSH3 ciblant différents sites de MSH3, ainsi quele recrutement de d'autres protéines essentielles à l'instabilité des triplets de CTG, comme MSH2, MSH6.

DISCUSSION ET PERSPECTIVES
La DM1 fait partie des dizaines de pathologies humaines dues à l'instabilité de triplets répétés instables. Dans la DM1, l'instabilité des triplets CTG est généralement biaisée vers les expansions au fil des générations et au cours de la vie des patients, en même temps que s'aggravent les symptômes (Hunter *et al.* 1992, Overend *et al.* 2019). Ainsi, l'instabilité des triplets répétés est un facteur clef de la pathogénèse des maladies à TNRs. Donc, une meilleure compréhension des mécanismes de l'instabilité des triplets est primordiale pour mieux appréhender les pistes diagnostiques.

De nombreuses études dans différents modèles murins, bactériens, de levures ou encore en cellules humaines et dans les familles de patients ont permis d'identifier de nombreux éléments modulateurs de l'instabilité. Parmi ces acteurs modulateurs, il a été identifié chez certaines familles de patients DM1 des interruptions CCG, CGG et CTC au sein des répétitions de CTG. Ces interruptions sont liées à une instabilité intergénérationnelle et somatique inhabituelles, avec une amplitude d'expansions diminuée et des symptômes moins sévères (voir (Santoro *et al.* 2017) pour revue et (Cumming *et al.* 2018, Cumming *et al.* 2019)). Cependant, les mécanismes par lesquels les interruptions influenceraient l'instabilité restent peu connus.

Dans ce contexte, mon travail a permis d'identifier un nouveau type d'interruption (une unique interruption CAG) et de démontrer que cette unique interruption est suffisante pour réduire l'instabilité des triplets CTG dans un modèle cellulaire humain. De plus, mes travaux apportent également de nouvelles données sur les mécanismes par lesquels les interruptions peuvent influencer l'instabilité des triplets répétés.

I- FAMILLES DM1 ATYPIQUES ET INTERRUPTIONS

Les interruptions dans les répétitions de CTG impliquées dans la DM1 ont été pour la première fois décrites il y a une dizaine d'années, avec des motifs variables de CTC et CCG en 3' des répétitions CTG (Musova *et al.* 2009). Dès lors, les interruptions ont été associées à d'inhabituelles contractions intergénérationnelles de triplets répétés ainsi qu'à un phénotype moins marqué et retardé. Depuis, ces interruptions ont été retrouvées dans plusieurs études familiales et il a été démontré que d'autres motifs à base d'interruptions GGC en 3' des répétitions et CCG en 5' pouvaient également survenir dans la DM1 (Braida *et al.* 2010, Santoro *et al.* 2013). Des interruptions CCG ont également été observées dans de rares cas d'allèle *DMPK* sain de 37 et 34 répétitions (P.Leeflang *et al.* 1995, Cumming *et al.* 2019). A noter que l'observation d'interruptions dans les allèles normaux est beaucoup plus fréquente dans les *loci* de FRAXA, FRDA, SCA1 et SCA2, où seules les répétitions ininterrompues sont pathologiques et instables. Ensembles, ces données suggèrent un effet stabilisateur des interruptions au sein des répétitions de triplets.

Notre travail apporte de nouvelles données sur ces interruptions, avec la description de l'existence d'un autre type d'interruption : une unique interruption CAG en 5' des répétitions CTG, stable au cours des transmissions. La famille DM1 porteuse de cette interruption présente une l'instabilité intergénérationnelle biaisée vers les contractions sur 5 transmissions maternelles et ne présente aucun phénomène d'anticipation. Or, dans le cas général, il serait habituellement attendu des transmissions de larges expansions suite à ces successions de transmissions maternelles, ce qui aboutirait à l'apparition de la forme congénitale (Brunner *et al.* 1993). De plus, l'instabilité somatique chez cette famille est d'amplitude réduite par rapport à des individus DM1 contrôles (à répétitions pures), mais aucun biais vers les contractions n'a été observé dans le sang. Ces résultats suggèrent que les interruptions, même uniques, peuvent impacter l'instabilité des triplets répétés, et que les mécanismes par lesquels ces interruptions interviennent diffèrent entre l'instabilité somatique et intergénérationnelle.

Cette hypothèse est d'autant plus confortée que nous avons éliminé d'autres facteurs connus pour influencer l'instabilité des triplets répétés :

- (1) la famille ne présentait pas de marque de méthylation différente de contrôles DM1 aux abords de DMPK; or, les différences de méthylation sont connues pour changer la dynamique mutationnelle des répétitions CTG dans des modèles murins SCA1 et SCA7 (Dion et al. 2008, Libby et al. 2008)
- (2) la transcription bidirectionnelle de *DMPK* mutant y est préservée. Une absence de la transcription de *DMPK* ne peut donc pas ici expliquer la stabilisation des triplets CTG dans cette famille (Nakamori *et al.* 2010)

Il n'est cependant pas à exclure qu'un autre facteur en *trans* des répétitions CTG puisse être au minimum co-acteur de la stabilisation des répétitions chez les familles interrompues comme un variant dans les protéines du MMR. En effet, il a récemment été observé que le polymorphisme de *MSH3* chez les patients influence l'instabilité des triplets (Flower *et al.* 2019). Chez la famille porteuse de l'interruption CAG, aucun polymorphisme marquant de MSH2 et MSH3 n'a été révélé lors d'analyse de co-ségrégation. Mais il pourrait être informatif de vérifier s'il n'y a pas chez ces familles la présence de d'autres polymorphismes pouvant influencer d'autres acteurs du MMR ou du traitement d'autres facteurs modulateurs de l'instabilité (R-loops, structures secondaires par exemples). Par ailleurs, il pourra aussi être informatif d'analyser le statut d'autres régions répétées polymorphes du génome comme d'autres séquences microsatellites, les séquences répétées Alu ou les séquences télomériques interstitielles (Mondello *et al.* 2000, Ade *et al.* 2013, Kim *et al.* 2014): est-ce que par exemple, chez ces familles à interruptions, les séquences répétées du génome en général sont en moyenne plus courtes que dans la population générale ?

L'obtention de différents tissus provenant de cette famille pourrait également être informatif à propos de l'impact des interruptions sur la tissu-spécificité de l'instabilité somatique. Cependant, la multiplication des prélèvements sur patients est difficilement envisageable. Pour pallier à cela, le développement de modèles cellulaires iPS isogéniques dérivés de ces familles et ensuite redifférenciés en différents types cellulaires (en myoblastes ou neurones par exemple) pourrait être un bon compromis.

Les différentes études dans les cohortes de familles DM1 montrent que les interruptions ne sont pas des phénomènes si isolés qu'on aurait pu le croire, entre 3 à 7% dans différentes cohortes de patients DM1 étudiées (Braida *et al.* 2010, Santoro *et al.* 2013, Botta *et al.* 2017, Cumming *et al.* 2018, Cumming *et al.* 2019). À la lumière de cette fréquence d'apparition non négligeable, et sachant que les interruptions peuvent interférer avec les techniques de diagnostic (Dryland *et al.* 2013, Santoro *et al.* 2017) le développement de techniques de TP-PCR en 5' et 3' et spécifique des interruptions, ainsi que la démocratisation de techniques de séquençage davantage résolutives que le séquençage Sanger comme par exemple le PacBio (Cumming *et al.* 2018) semble devenir primordial pour affiner le diagnostic. De plus, la symptomatologie de la DM1 est très variable d'un individu à un autre, et jusqu'à présent les interruptions semblent être liées à une pénétrance réduite des pathologies à TNRs (voir chapitre III.3.1). Il apparaît nécessaire d'en comprendre la survenue ainsi que les mécanismes par lesquels les différents types d'interruptions interviennent afin de pouvoir proposer un pronostic et un conseil génétique davantage personnalisés aux patients, voire d'envisager des thérapies mimant l'effet d'interruptions.

Enfin, il me semble important de souligner ici que nos travaux présentés dans ce chapitre, ainsi que de bon nombre d'autres données récentes sur les interruptions n'auraient jamais pu être réalisées sans le développement de banques de données de patients telles que le DM-Scope et le réseau OPTIMISTIC. Ces réseaux qui compilent les données génotypiques et phénotypiques des patients permettent et accélèrent grandement les protocoles de comparaisons entre patients atypiques et patients DM1 contrôles. Le développement de ce type d'approches avec la mise en place d'échanges raisonnés avec les laboratoires de recherche représentent un fort potentiel pour l'accessibilité à de nouvelles pistes de recherche.

II- LA FLASH-SMALL-POOL PCR

La Small-Pool PCR est la technique privilégiée pour l'analyse de l'instabilité somatique des triplets répétés. Cette technique permet d'amplifier un faible nombre de molécules d'ADN et, jusqu'à présent, nécessitait un Southern Blot. Bien que cette technique présente une excellente résolution pour l'analyse de l'instabilité, la Small-Pool PCR demeurait relativement chère et fastidieuse à effectuer, notamment à cause de l'achat de membranes, de films autoradiographiques, d'anticorps ou la nécessité de disposer d'un espace dédié à la radioactivité. De plus, les protocoles précédents demandaient au minimum 3 jours d'expérimentation sans qu'il ne soit possible de contrôler la validité de l'expérience en cours avant sa révélation finale.

La Flash-Small-Pool-PCR (FSP-PCR) repose toujours sur l'amplification d'un faible nombre de molécules d'ADN, mais avec l'un des 2 primers marqué en fluorescence et la résolution du produit de PCR par migration sur gel d'agarose dénaturant. Le gel est ensuite directement placé sur un appareil capable de détecter et capturer et capturer le signal fluorescent de l'amorce marquée. Le protocole peut encore être réduit en temps en faisant migrer le gel à un voltage plus important à 4°C, mais cela nécessite de changer régulièrement le tampon de migration pour éviter la surchauffe.

Le protocole proposé ici présente différents avantages. Tout d'abord, il permet de réduire le temps d'expérimentation à une journée et réduit le nombre d'étapes pouvant insérer un biais expérimental, ce qui, en cas d'analyse d'un grand nombre d'échantillons peut représenter un gain de temps substantiel. Par ailleurs, la FSP-PCR ne nécessitant pas d'optimisation de PCR, sa mise ne place ne demande que l'achat d'une seule amorce marquées en fluorescence en plus des réactifs utilisés en routine pour la PCR. De plus, ce protocole ne nécessite plus l'achat de membranes, de films radiographique, d'anticorps, de sondes LNA ou radioactives. Cependant, la principale limitation matérielle nécessaire à la mise en place de ce protocole est l'accessibilité à un appareil capable de lire la fluorescence proche Infra-rouge sur gel d'agarose. Par ailleurs, cette technique n'améliore pas l'amplification en

elle-même : les tailles de triplets non amplifiées auparavant ne seront pas détectables non plus avec cette technique. Les grandes tailles de répétitions peuvent en effet être très difficiles à amplifier, et le protocole de Flash-Small-Pool ne permet d'améliorer cet écueil. Ainsi, pour la détection de très grandes tailles de triplets, une hybridation avec une sonde spécifique des triplets reste à préconiser. Par ailleurs ce protocole n'améliore pas non plus le biais de PCR qui peut survenir entre l'amplification favorisée de l'allèle sain et celle de l'allèle mutant. Au contraire, l'allèle sain donne généralement un signal beaucoup plus fort pouvant parfois gêner la bonne détection de la mosaïque somatique en cas de contraction importantes ou de petites tailles d'expansions.

L'analyse de l'instabilité des triplets répétés représente une étape clef dans la compréhension des pathologies à TNRs. La démocratisation et facilitation de techniques comme la Small-Pool-PCR représentent donc une avancée intéressante pour la poursuite de la recherche sur les mécanismes de l'instabilité des triplets impliqués dans les pathologies humaines.

III- ROLE DES INTERRUPTIONS DANS L'INSTABILITE DES TRIPLETS CTG

Des études chez les patients ont déjà grandement suggéré le rôle stabilisateur des interruptions dans la DM1 (voir (Santoro et al. 2017) pour revue, (Cumming et al. 2018, Tomé et al. 2018, Cumming et al. 2019). Cependant, la potentielle existence d'autres facteurs en *trans* pouvant influencer l'instabilité (polymorphismes dans les acteurs du MMR comme *MSH3* par exemple), limitait les conclusions sur les effets imputables aux interruptions dans ces analyses chez les patients. Il restait à démontrer que les interruptions soient suffisantes pour moduler l'instabilité des triplets. En effet, les interruptions auraient également pu n'être qu'un biomarqueur témoignant d'un mécanisme agissant en amont de leur survenue.

III.1. UNE UNIQUE INTERRUPTION CAG DIMINUE L'INSTABILITE DES TRIPLETS CTG

Mon travail a permis de démontrer qu'une seule interruption CAG en 5' des répétitions était capable de limiter la taille des expansions dans un modèle de cellules humaines porteuses de 123 répétitions CTG, et cela jusqu'à au moins 42 semaines en culture. Il est intéressant de constater que malgré la variabilité clonale, que l'intégralité des clones à répétitions pures ont des expansions que larges que les clones à répétitions interrompues (voir article 3 figure 3). Ainsi, notre modèle reproduit ce qui était observé chez la famille porteuse de l'interruption CAG : une instabilité somatique à amplitude réduite comparée aux répétitions pures. Ce modèle est donc pertinent pour l'analyse de l'impact des interruptions sur l'instabilité des triplets de CTG. Il sera informatif de créer d'autres lignées porteuses de répétitions de CTG interrompues par différentes interruptions (CCG, CGG et CTC) pour comparer leur

influence sur l'instabilité des triplets CTG : limitent-elles toutes l'instabilité ? Est-ce que les interruptions protègent toutes avec la même efficacité contre les grandes expansions de triplets ?

III.2. UNE UNIQUE INTERRUPTION CAG INFLUENCE LA TRANSCRIPTION DE *DMPK*

Au-delà de la démonstration de l'impact direct d'une unique interruption sur l'instabilité des triplets CTG, ce modèle apporte des précisions sur l'influence de l'interruption sur la transcription de DMPK, connue pour stimuler l'instabilité des triplets (Nakamori et al. 2010). Plus particulièrement, la production d'R-loops et le niveau de production du transcrit antisens rCAG de DMPK mutant sont réduits en cas d'interruption CAG en 5' des interruptions. De plus, il a été observé une corrélation entre le niveau de transcription antisens rCAG et la formation d'R-loop. Il est possible que comme la transcription de rCAG est réduite en cas d'interruption, il y a moins de transcrits disponible pour former les R-loops, ce qui expliquerait la diminution de la production d'R-loops dans les clones à interruption CAG. Il a cependant été démontré que que les R-loops favorisaient la transcription antisens de façon générale dans le génome (Skourti-Stathaki et al. 2014). Ainsi, une autre théorie serait que l'interruption CAG déstabilise la production de R-loops, ce qui réduirait la stimulation de la transcritption antisens de rCAG dans notre modèle. Afin de conclure sur la question, il sera nécessaire quantifier spécifiquement les transcrits rCAG et rCUG impliqués dans les R-loops. Les R-loops, en sont propices à l'instabilité des triplets : pause de la machinerie de transcription, formation de structures secondaires, augmentation de la probabilité de collision entre les machineries de transcription en sens et en antisens, ou des machineries de transcription avec celles de la réplication (voir Introduction chapitre III.4.3.4 pour plus de détails). La diminution de la présence d'R-loops dans le cas de répétitions interrompues diminue la probabilité de la survenue de ces évènements favorisant, ce qui pourrait expliquer au moins en partie la stabilisation des triplets CTG en cas d'interruption. Il serait intéressant de tester l'impact d'autres natures interruptions sur la transcription. Il sera aussi nécessaire de valider ce résultat avec des promoteurs de la transcription plus semblables à la séquence DMPK in vivo. En effet, dans notre modèle, la transcription est induite de façon constitutive par des promoteurs ROSA26 et CMV ne correspondant pas aux promoteurs de la transcription de DMPK natif. Une telle expérience permettrait de déterminer plus précisément dans quelle mesure l'interruption impacte la transcription de DMPK.

En conclusion, cette étude démontre qu'une unique interruption CAG en 5' des répétitions CTG est capable de diminuer l'amplitude de l'instabilité somatique, la production de transcrits rCAG mutants et la formation d'R-loops. L'une des principales limitations du modèle cellulaire utilisé dans cette étude est l'insertion aléatoire de la région 3'UTR de *DMPK* mutant. Or, il a été démontré que l'environnement dans lequel se trouve les répétitions est

primordial pour l'instabilité des triplets. Pour construire ce modèle, je me suis appuyée sur l'action de la Phi intégrase qui compte une centaine de site d'insertion sans réelle préférences. Il est ainsi très probable que chaque clone présente un site d'insertion différent. Ceci peut notamment une part de la variabilité inter-clonale observée. Par ailleurs, notre modèle présente un niveau de transcrit antisens supérieur à celui du transcrit sens : ce qui s'oppose à ce qui est observé chez les patients (Gudde *et al.* 2017). Ceci peut être la conséquence d'une efficacité différente entre facteurs de transcription utilisés pour induire l'expression bidirectionnelle de *DMPK* 3'UTR mutant. En effet, une étude dans des cellules ES de souris a démontré que l'activité de ROSA26 (promoteur de la transcription du brin rCAG antisens) était plus importante que celle du promoteur CMV (utilisé pour la transcription du brin rCUG sens), lorsque CMV était inséré sur le *locus* de ROSA26 (Chen *et al.* 2011).

Ainsi, l'idéal pour la poursuite de l'analyse de l'impact des interruptions sur l'instabilité des triplets sera de développer un modèle cellulaire isogénique avec une insertion ciblée des répétitions de CTG pures et interruptions dans le *locus DMPK* par technologie Crispr/Cas9. Le projet est déjà amorcé dans le laboratoire.

III.3. LES INTERRUPTIONS MODIFIENT LA STRUCTURE SECONDAIRE DES REPETITIONS DE CTG

J'ai démontré dans des expériences à base d'amplicons de PCR, que l'interruption CAG modifie la structure secondaire des répétitions CTG. Les expériences de digestions enzymatiques mettent en évidence que les répétitions pures et interrompues semblent toutesdeux former des structures simples brins et des mismatches. La différence de structures entre les répétitions pures et interrompues résiderait donc soit dans la stabilité des structures, soit dans leur localisation et leur nombre le long des répétitions. En effet, une étude d'analyse de point de fusion a démontré que les répétitions CTG et CAG formaient plus fréquemment de multiples hairpins de tailles modestes qu'une seule grande hairpin (Petruska *et al.* 1996) Pour pouvoir statuer sur l'impact des interruptions sur les structures secondaires, il pourra-être intéressant de localiser les structures simples brin avec des expériences à base de traitement aux bisulfites en conditions natives couplé à un séquençage par exemple. On peut également suggérer une analyse de la stabilité des structures par des expériences de dichroïsme circulaire ou de résonnance magnétique ayant déjà prouvé leur efficacité dans la détermination de structures secondaires impliquant des triplets et des mismatches (Gacy *et al.* 1995, Teng *et al.* 2018).

III.4. LES INTERRUPTIONS MODIFIENT LE RECRUTEMENT DE PROTEINES NUCLEAIRES SUR LES REPETITIONS DE CTG

Il a été démontré que l'interaction des répétitions CTG avec différents acteurs du métabolisme de l'ADN était primordial dans l'instabilité des triplets (voir Introduction chapitre III.4.). Au cours de ma thèse, j'ai démontré qu'une unique interruption CAG pouvait influencer le recrutement de protéines nucléaires sur les répétitions CTG par EMSA. Mes résultats préliminaires tendent à démontrer que les interruptions CCG seraient également capables de modifier ce recrutement, mais de manière différente de l'interruption CAG. Ce résultat suggère donc que selon la nature des interruptions, leur impact sur le métabolisme de l'ADN peut être différent : les conséquences des interruptions sur l'instabilité pourraient donc être différentes aussi, ce qui vient souligner l'importance du développement de modèles capables de révéler l'impact de différentes interruptions sur l'instabilité des triplets CTG en fonction de leur nature.

Le MMR, et tout particulièrement *MSH3* a un rôle crucial dans l'instabilité des triplets CTG (voir chapitre III.4.2.3). C'est pourquoi j'ai testé le recrutement de MSH3 sur les séquences pures et interrompues par Supershift, et j'ai démontré que MSH3 était recruté sur les répétitions pures et interrompues. Cependant, en cas de répétitions pure, un complexe n'est pas à l'anticorps anti-MSH3 contrairement à un complexe présentant le même profil de migration sur gel en cas de répétitions interrompues. Ce qui suggère que l'interruption CAG modifie la disponibilité de MSH3 pour l'anticorps, soit en changeant son interaction avec les répétitions et rendant l'interaction de MSH3 avec l'anticorps utilisé possible, soit en modulant le recrutement d'un potentiel co-facteur qui masquait le site de reconnaissance de MSH3 pour cet anticorps en cas de répétitions pures (voir article 3 figure 2).

Afin de mieux comprendre comment est-ce que les interruptions peuvent influencer le recrutement de protéines nucléaires sur les répétitions de CTG, il sera nécessaire de tester d'autres anticorps ciblant différents sites de MSH3 et en différentes quantités. Par ailleurs, il sera intéressant de tester également par supershift le recrutement d'autres facteurs essentiels à l'instabilité des triplets comme MSH2 et MSH6 (en cours au laboratoire). Ces informations pourront être obtenues par superhifts ou par spectrophotométrie de masse. Il pourra aussi être très informatif de réaliser des expériences de co-immunoprécipitation visant ces protéines candidates dans le modèle cellulaire présenté dans ces travaux, ainsi que dans le modèle de cellules isogéniques en devenir. Ainsi, il sera possible d'analyser les facteurs recrutés différemment en fonction de la nature des répétitions, mais aussi d'évaluer l'affinité de ces facteur en fonction de la nature des triplets CTG, en couplant l'expérience de co-immunoprécipitation avec une qPCR ciblant *DMPK*.

V. CONCLUSION GENERALE

Pour conclure, les travaux réalisés au cours de ma thèse ont permis d'identifier une nouvelle interruption CAG unique en 5' de la répétition de triplets CTG chez une famille de patients DM1 atypique (instabilité intergénérationnelle biaisée vers les contractions, pas de phénomène d'anticipation, amplitude de la mosaïque somatique réduite). J'ai adapté un modèle cellulaire humain qui m'a permis de démontrer qu'une unique interruption CAG en 5' des répétitions de CTG est capable de réduire l'amplitude de la mosaïque somatique. Ce modèle a également permis de démontrer que la présence de cette interruption diminue le niveau de transcription antisens (transcrit rCAG), en corrélation avec une diminution de la formation d'R-loops. Or, la production du transcrit antisens rCAG ainsi que des R-loops sont des facteurs connus pour favoriser l'instabilité des triplets. J'ai également démontré que les interruptions induisent différentes structures secondaires qu'en cas de répétitions pures, modifiant ainsi potentiellement l'accessibilité des répétitions aux acteurs du métabolisme de l'ADN. Cette suggestion a été confortée par EMSA et supershift, expériences ayant démontré que les interruptions modulent le recrutement de protéines nucléaires sur les répétitions de CTG, notamment de MSH3. L'ensemble de ces travaux proposent ainsi des pistes sur les mécanismes par lesquels les interruptions peuvent influencer l'instabilité des triplets répétés : la formation de structures secondaire, le recrutement différent de protéines sur les répétitions de CTG dont du MMR, la transcription bidirectionnelle et la formation de R-loops (voir figure 25).

Ce travail a prêté une attention particulière à l'unique interruption CAG et gagnera à être reproduit sur d'autres types d'interruptions identifiées jusqu'ici. Notamment les interruption CCG qui ont été plus fréquemment identifiées dans les familles de patients DM1. Enfin, le développement de lignées de cellules isogéniques à insertion ciblée des répétitions de CTG pures et interrompues seront des modèles d'autant plus représentatifs de ce qui peut se produire chez les patients, et seront de précieux outils pour davantage comprendre l'impact des interruptions sur l'instabilité des triplets CTG.



Figure 24: Modèle intégratif de l'impact de l'interruption CAG dans l'instabilité des triplets CTG.

Les travaux présentés au cours de cette thèse a permis de démontrer que la nature des répétitions (pure ou interrompue) influence la structure secondaire des répétitions, le recrutement de protéines nucléaires (notamment MSH3), et modifient le niveau de transcription du transcrit rCAG et la formation d'R-loops qui sont corrélés entreeux (flèches pleines). Il est raisonnable de supposer (flèches en pointillés) que les variations de structures secondaires et de recrutement de protéines sur les répétitions à cause d'interruption puissent influencer la transcription, la réplication et la réparation de l'ADN au niveau des répétitions. Par ailleurs, la modulation du niveau de R-loops (diminution en cas d'interruptions CAG) peut également influencer la réplication et la réparation de l'ADN mécanismes connus pour moduler d'instabilité des TNRs.

Une meilleure compréhension de l'instabilité des triplets répétés est primordiale pour mieux appréhender les maladies à TNRs et développer des stratégies thérapeutiques adaptées. La multiplication ces dernières années des observations d'interruptions liées à des patients à l'instabilité de triplets répétés atypiques et à l'amoindrissement des symptômes soulignent l'importance des interruptions dans le paysage de l'instabilité des TNRs, et, surtout, dans leur stabilisation voire contractions, phénomènes dont les mécanismes sont à ce jour peu connus.

BIBLIOGRAPHIE

Abu-Baker, A., C. Messaed, J. Laganiere, C. Gaspar, B. Brais and G. A. Rouleau (2003). "Involvement of the ubiquitin-proteasome pathway and molecular chaperones in oculopharyngeal muscular dystrophy." <u>Human molecular genetics</u> **12**(20): 2609-2623 %@ 1460-2083.

Ade, C., A. M. Roy-Engel and P. L. Deininger (2013). "Alu elements: an intrinsic source of human genome instability." <u>Current Opinion in Virology</u> **3**(6): 639-645.

Adihe Lokanga, R., X.-N. Zhao, A. Entezam and K. Usdin (2014). "X inactivation plays a major role in the gender bias in somatic expansion in a mouse model of the fragile X-related disorders: implications for the mechanism of repeat expansion." <u>Human molecular genetics</u> **23**(18): 4985-4994.

Aeschbach, L. and V. Dion (2017). "Minimizing carry-over PCR contamination in expanded CAG/CTG repeat instability applications." <u>Scientific Reports</u> **7**(1): 18026.

Agrawal, S. and A. R. J. P. o. Ganley (2018). "The conservation landscape of the human ribosomal RNA gene repeats." <u>PloS one</u> **13**(12): e0207531.

Aguilera, A. and T. García-Muse (2012). "R Loops: From Transcription Byproducts to Threats to Genome Stability." <u>Molecular Cell</u> **46**(2): 115-124.

Al-Mahdawi, S., R. M. Pinto, P. Ruddle, C. Carroll, Z. Webster and M. Pook (2004). "GAA repeat instability in Friedreich ataxia YAC transgenic mice." <u>Genomics</u> **84**(2): 301-310.

Albrecht, A. and S. Mundlos (2005). "The other trinucleotide repeat: polyalanine expansion disorders." <u>Current opinion in genetics & development</u> **15**(3): 285-293 %@ 0959-0437X.

Albrecht, A. N., U. Kornak, A. Böddrich, K. Süring, P. N. Robinson, A. C. Stiege, R. Lurz, S. Stricker, E. E. Wanker and S. Mundlos (2004). "A molecular pathogenesis for transcription factor associated poly-alanine tract expansions." <u>Human Molecular Genetics</u> **13**(20): 2351-2359.

Amiel, J., D. Trochet, M. Clément-Ziza, A. Munnich and S. Lyonnet (2004). "Polyalanine expansions in human." <u>Human Molecular Genetics</u> **13**(suppl_2): R235-R243.

Arrasate, M., S. Mitra, E. S. Schweitzer, M. R. Segal and S. Finkbeiner (2004). "Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death." <u>nature</u> **431**(7010): 805 %@ 1476-4687.

Ashizawa, T., M. Anvret, M. Baiget, J. M. Barcelo, H. Brunner, A. M. Cobo, B. Dallapiccola, R. G. Fenwick, Jr., U. Grandell, H. Harley and et al. (1994). "Characteristics of intergenerational contractions of the CTG repeat in myotonic dystrophy." <u>Am J Hum Genet</u> **54**(3): 414-423.

Ashizawa, T., J. R. Dubel and Y. Harati (1993). "Somatic instability of CTG repeat in myotonic dystrophy." <u>Neurology</u> **43**(12): 2674-2674.

Babačić, H., A. Mehta, O. Merkel and B. Schoser (2019). "CRISPR-cas gene-editing as plausible treatment of neuromuscular and nucleotide-repeat-expansion diseases: A systematic review." <u>PloS one</u> **14**(2): e0212198-e0212198.

Barbé, L., S. Lanni, A. López-Castel, S. Franck, C. Spits, K. Keymolen, S. Seneca, S. Tomé, I. Miron, J. Letourneau, M. Liang, S. Choufani, R. Weksberg, M. D. Wilson, Z. Sedlacek, C. Gagnon, Z. Musova, D. Chitayat, P. Shannon, J. Mathieu, K. Sermon and C. E. Pearson (2017). "CpG Methylation, a Parent-of-Origin Effect for Maternal-Biased Transmission of Congenital Myotonic Dystrophy." <u>American journal of human genetics</u> **100**(3): 488-505.

Barcelo, J. M., M. Pluscauskas, A. E. MacKenzie, C. Tsilfidis, M. Narang and R. G. Korneluk (1994). "Additive influence of maternal and offspring DM-kinase gene CTG repeat lengths in the genesis of congenital myotonic dystrophy." <u>Am J Hum Genet</u> **54**(6): 1124-1125.

Barthelemy, J., H. Hanenberg and M. Leffak (2016). "FANCJ is essential to maintain microsatellite structure genome-wide during replication stress." <u>Nucleic Acids Res</u> **44**(14): 6803-6816.

Bassez, G. (2019). <u>The iDM-Scope Registry: An International Framework to Support Myotonic</u> <u>Dystrophy Translational Research</u>. IDMC12, Gothenburg, Sweden.

Bassez, G., E. Audureau, J.-Y. Hogrel, R. Arrouasse, S. Baghdoyan, H. Bhugaloo, M.-L. Gourlay-Chu, P. Le Corvoisier and M. Peschanski (2018). "Improved mobility with metformin in patients with myotonic dystrophy type 1: a randomized controlled trial." <u>Brain</u> **141**(10): 2855-2865.

Batra, R., K. Charizanis, M. Manchanda, A. Mohan, M. Li, D. J. Finn, M. Goodwin, C. Zhang, K. Sobczak and C. A. Thornton (2014). "Loss of MBNL leads to disruption of developmentally regulated alternative polyadenylation in RNA-mediated disease." <u>Molecular cell</u> **56**(2): 311-322.

Batra, R., D. A. Nelles, E. Pirie, S. M. Blue, R. J. Marina, H. Wang, I. A. Chaim, J. D. Thomas, N. Zhang, V. Nguyen, S. Aigner, S. Markmiller, G. Xia, K. D. Corbett, M. S. Swanson and G. W. Yeo (2017). "Elimination of Toxic Microsatellite Repeat Expansion RNA by RNA-Targeting Cas9." <u>Cell</u> **170**(5): 899-912.e810.

Bäuerlein, F. J. B., I. Saha, A. Mishra, M. Kalemanov, A. Martínez-Sánchez, R. Klein, I. Dudanova, M. S. Hipp, F. U. Hartl and W. Baumeister (2017). "In situ architecture and cellular interactions of PolyQ inclusions." <u>Cell</u> **171**(1): 179-187. e110 %@ 0092-8674.

Belotserkovskii, B. P., S. Tornaletti, A. D. D'Souza and P. C. Hanawalt (2018). "R-loop generation during transcription: Formation, processing and cellular outcomes." <u>DNA Repair</u> (<u>Amst</u>) **71**: 69-81.

Berul, C. I., C. T. Maguire, M. J. Aronovitz, J. Greenwood, C. Miller, J. Gehrmann, D. Housman, M. E. Mendelsohn and S. Reddy (1999). "DMPK dosage alterations result in atrioventricular conduction abnormalities in a mouse myotonic dystrophy model." <u>The Journal of clinical investigation</u> **103**(4): R1-R7.

Bettencourt, C., D. Hensman-Moss, M. Flower, S. Wiethoff, A. Brice, C. Goizet, G. Stevanin, G. Koutsis, G. Karadima, M. Panas, P. Yescas-Gómez, L. E. García-Velázquez, M. E. Alonso-Vilatela, M. Lima, M. Raposo, B. Traynor, M. Sweeney, N. Wood, P. Giunti, S. Network, A. Durr, P. Holmans, H. Houlden, S. J. Tabrizi and L. Jones (2016). "DNA repair pathways underlie a common genetic mechanism modulating onset in polyglutamine diseases." <u>Annals of neurology</u> **79**(6): 983-990.

Bird, T. D. (2018). "Myotonic Dystrophy Type 1."

Bisset, D. R., E. A. Stepniak-Konieczna, M. Zavaljevski, J. Wei, G. T. Carter, M. D. Weiss and J. R. Chamberlain (2015). "Therapeutic impact of systemic AAV-mediated RNA interference in a mouse model of myotonic dystrophy." <u>Human molecular genetics</u> **24**(17): 4971-4983.

Bochkareva, A., Y. Yuzenkova, V. R. Tadigotla and N. Zenkin (2012). "Factor-independent transcription pausing caused by recognition of the RNA–DNA hybrid sequence." <u>The EMBO</u> journal **31**(3): 630-639.

Botta, A., G. Rossi, M. Marcaurelio, L. Fontana, M. R. D'Apice, F. Brancati, R. Massa, D. G Monckton, F. Sangiuolo and G. Novelli (2017). "Identification and characterization of 5' CCG interruptions in complex DMPK expanded alleles." <u>European journal of human genetics : EJHG</u> **25**(2): 257-261.

Bourn, R. L., I. De Biase, R. M. Pinto, C. Sandi, S. Al-Mahdawi, M. A. Pook and S. I. Bidichandani (2012). "Pms2 suppresses large expansions of the (GAA·TTC)n sequence in neuronal tissues." <u>PloS one</u> **7**(10): e47085-e47085.

Boutell, J. M., P. Thomas, J. W. Neal, V. J. Weston, J. Duce, P. S. Harper and A. Lesley Jones (1999). "Aberrant interactions of transcriptional repressor proteins with the Huntington's disease gene product, huntingtin." <u>Human molecular genetics</u> **8**(9): 1647-1655 %@ 1460-2083.

Braida, C., R. K. A. Stefanatos, B. Adam, N. Mahajan, H. J. M. Smeets, F. Niel, C. Goizet, B. Arveiler, M. Koenig, C. Lagier-Tourenne, J.-L. Mandel, C. G. Faber, C. E. M. de Die-Smulders, F. Spaans and D. G. Monckton (2010). "Variant CCG and GGC repeats within the CTG expansion dramatically modify mutational dynamics and likely contribute toward unusual symptoms in some myotonic dystrophy type 1 patients." <u>Human Molecular Genetics</u> **19**(8): 1399-1412.

Braz, S. O., J. Acquaire, G. Gourdon and M. Gomes-Pereira (2018). "Of mice and men: advances in the understanding of neuromuscular aspects of Myotonic Dystrophy." <u>Frontiers in neurology</u> **9**: 519.

Brock, G. J., N. H. Anderson and D. G. Monckton (1999). "Cis-acting modifiers of expanded CAG/CTG triplet repeat expandability: associations with flanking GC content and proximity to CpG islands." <u>Hum Mol Genet</u> **8**(6): 1061-1067.

Brook, J. D., M. E. McCurrach, H. G. Harley, A. J. Buckler, D. Church, H. Aburatani, K. Hunter, V. P. Stanton, J.-P. Thirion and T. J. C. Hudson (1992). "Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member." <u>Cell</u> **68**(4): 799-808.

Brouwer, J. R., A. Huguet, A. Nicole, A. Munnich and G. Gourdon (2013). "Transcriptionally Repressive Chromatin Remodelling and CpG Methylation in the Presence of Expanded CTG-Repeats at the DM1 Locus." Journal of nucleic acids **2013**: 567435-567435.

Brunner, H., H. Brüggenwirth, W. Nillesen, G. Jansen, B. Hamel, R. Hoppe, C. de Die, C. Höweler, B. Van Oost and B. Wieringa (1993). "Influence of sex of the transmitting parent as well as of parental allele site on the CTG expansion in myotonic dystrophy (DM)." <u>American</u> journal of human genetics **53**(5): 1016.

Budworth, H., F. R. Harris, P. Williams, D. Y. Lee, A. Holt, J. Pahnke, B. Szczesny, K. Acevedo-Torres, S. Ayala-Peña and C. T. McMurray (2015). "Suppression of Somatic Expansion Delays the Onset of Pathophysiology in a Mouse Model of Huntington's Disease." <u>PLOS Genetics</u> **11**(8): e1005267.

Budworth, H. and C. T. McMurray (2013). "Bidirectional transcription of trinucleotide repeats: roles for excision repair." <u>DNA repair</u> **12**(8): 672-684.

Bürk, K. (2017). "Friedreich Ataxia: current status and future prospects." <u>Cerebellum & Ataxias</u> **4**(1): 4.

Burman, R. W., B. W. Popovich, P. B. Jacky and M. S. Turker (1999). "Fully expanded FMR1 CGG repeats exhibit a length- and differentiation-dependent instability in cell hybrids that is independent of DNA methylation." <u>Hum Mol Genet</u> **8**(12): 2293-2302.

Caburet, S., A. Demarez, L. Moumné, M. Fellous, E. De Baere and R. A. Veitia (2004). "A recurrent polyalanine expansion in the transcription factor FOXL2 induces extensive nuclear and cytoplasmic protein aggregation." Journal of Medical Genetics **41**(12): 932.

Caillet-Boudin, M.-L., F. Fernandez-Gomez, H. Tran-Ladam, c.-m. dhaenens, L. Buee and N. Sergeant (2014). "Brain pathology in myotonic dystrophy: when tauopathy meets spliceopathy and RNAopathy." <u>Frontiers in Molecular Neuroscience</u> **6**(57).

Carrell, S. T., E. M. Carrell, D. Auerbach, S. K. Pandey, C. F. Bennett, R. T. Dirksen and C. A. Thornton (2016). "Dmpk gene deletion or antisense knockdown does not compromise cardiac or skeletal muscle function in mice." <u>Human molecular genetics</u> **25**(19): 4328-4338.

Carroll, M. S., G. Webster and D. E. Weese-Mayer (2017). Heart Rate Variability Measures From 24-Hour Holter Monitoring In Congenital Central Hypoventilation Syndrome (CCHS): Correlations Between Physiological Impairment And Polyalanine Repeat Length In The Paired-Line Homeobox 2B Gene. <u>A107. PEDIATRIC CARDIOVASCULAR DISEASE AND CRITICAL</u> <u>CARE</u>, American Thoracic Society: A2822-A2822 %@ 1073-2449X.

Cavalli-Sforza, L. L. and F. Cavalli-Sforza (2006). "The Great Human Diasporas: The History of Diversity and Evolution (Helix Books)."

Cayrou, C., P. Coulombe, A. Vigneron, S. Stanojcic, O. Ganier, I. Peiffer, E. Rivals, A. Puy, S. Laurent-Chabalier, R. Desprat and M. Méchali (2011). "Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features." <u>Genome research</u> **21**(9): 1438-1449.

Chan, N. L., J. Guo, T. Zhang, G. Mao, C. Hou, F. Yuan, J. Huang, Y. Zhang, J. Wu, L. Gu and G. M. Li (2013). "Coordinated processing of 3' slipped (CAG)n/(CTG)n hairpins by DNA polymerases beta and delta preferentially induces repeat expansions." <u>J Biol Chem</u> **288**(21): 15015-15022.

Chao, M. J., T. Gillis, R. S. Atwal, J. S. Mysore, J. Arjomand, D. Harold, P. Holmans, L. Jones, M. Orth, R. H. Myers, S. Kwak, V. C. Wheeler, M. E. MacDonald, J. F. Gusella and J.-M. Lee (2017). "Haplotype-based stratification of Huntington's disease." <u>European journal of human genetics : EJHG</u> **25**(11): 1202-1209.

Charlesworth, B., P. Sniegowski and W. Stephan (1994). "The evolutionary dynamics of repetitive DNA in eukaryotes." <u>Nature</u> **371**(6494): 215-220.

Charlet-B, N., R. S. Savkur, G. Singh, A. V. Philips, E. A. Grice and T. A. Cooper (2002). "Loss of the muscle-specific chloride channel in type 1 myotonic dystrophy due to misregulated alternative splicing." <u>Molecular cell</u> **10**(1): 45-53.

Chastain, P. D., 2nd, E. E. Eichler, S. Kang, D. L. Nelson, S. D. Levene and R. R. Sinden (1995). "Anomalous rapid electrophoretic mobility of DNA containing triplet repeats associated with human disease genes." <u>Biochemistry</u> **34**(49): 16125-16131.

Chau, A. and A. Kalsotra (2015). "Developmental insights into the pathology of and therapeutic strategies for DM1: back to the basics." <u>Developmental Dynamics</u> **244**(3): 377-390.

Chauhan, C., D. Dash, D. Grover, J. Rajamani and M. Mukerji (2002). "Origin and instability of GAA repeats: insights from Alu elements." <u>J Biomol Struct Dyn</u> **20**(2): 253-263.

Chédin, F. (2016). "Nascent Connections: R-Loops and Chromatin Patterning." <u>Trends in</u> genetics : TIG **32**(12): 828-838.

Chen, C.-m., J. Krohn, S. Bhattacharya and B. Davies (2011). "A comparison of exogenous promoter activity at the ROSA26 locus using a Φ iC31 integrase mediated cassette exchange approach in mouse ES cells." <u>PloS one</u> **6**(8): e23376-e23376.

Chen, C. M., Y. T. Hou, J. Y. Liu, Y. R. Wu, C. H. Lin, H. C. Fung, W. C. Hsu, Y. Hsu, S. H. Lee and H. M. J. A. J. o. M. G. P. B. N. G. Hsieh-Li (2009). "PPP2R2B CAG repeat length in the Han Chinese in Taiwan: association analyses in neurological and psychiatric disorders and potential functional implications." **150**(1): 124-129.

Chen, P. C., S. Dudley, W. Hagen, D. Dizon, L. Paxton, D. Reichow, S. R. Yoon, K. Yang, N. Arnheim, R. M. Liskay and S. M. Lipkin (2005). "Contributions by MutL homologues Mlh3 and Pms2 to DNA mismatch repair and tumor suppression in the mouse." <u>Cancer Res</u> **65**(19): 8662-8670.

Cherng, N., A. A. Shishkin, L. I. Schlager, R. H. Tuck, L. Sloan, R. Matera, P. S. Sarkar, T. Ashizawa, C. H. Freudenreich and S. M. Mirkin (2011). "Expansions, contractions, and fragility of the spinocerebellar ataxia type 10 pentanucleotide repeat in yeast." <u>Proceedings of the National Academy of Sciences</u> **108**(7): 2843-2848.

Chi, L. M. and S. L. Lam (2005). "Structural roles of CTG repeats in slippage expansion during DNA replication." <u>Nucleic Acids Res</u> **33**(5): 1604-1617.

Cho, D. H., C. P. Thienes, S. E. Mahoney, E. Analau, G. N. Filippova and S. J. Tapscott (2005). "Antisense transcription and heterochromatin at the DM1 CTG repeats are constrained by CTCF." <u>Mol Cell</u> **20**(3): 483-489.

Choudhry, S., M. Mukerji, A. K. Srivastava, S. Jain and S. K. Brahmachari (2001). "CAG repeat instability at SCA2 locus: anchoring CAA interruptions and linked single nucleotide polymorphisms." <u>Human molecular genetics</u> **10**(21): 2437-2446.

Chowdhury, M. R., S. Chauhan, A. Dabral, B. K. Thelma, N. Gupta and M. Kabra (2016). "Validation of Polymerase Chain Reaction–Based Assay to Detect Actual Number of CGG Repeats in FMR1 Gene in Indian Fragile X Syndrome Patients." <u>Journal of Child Neurology</u> **32**(4): 371-378.

Christofk, H. R., M. G. Vander Heiden, M. H. Harris, A. Ramanathan, R. E. Gerszten, R. Wei, M. D. Fleming, S. L. Schreiber and L. C. Cantley (2008). "The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth." <u>Nature</u> **452**(7184): 230.

Chung, M.-y., L. P. W. Ranum, L. A. Duvick, A. Servadio, H. Y. Zoghbi and H. T. Orr (1993). "Evidence for a mechanism predisposing to intergenerational CAG repeat instability in spinocerebellar ataxia type I." <u>Nature Genetics</u> **5**(3): 254-258.

Ciechanover, A. and P. Brundin (2003). "The ubiquitin proteasome system in neurodegenerative diseases: sometimes the chicken, sometimes the egg." <u>Neuron</u> **40**(2): 427-446 %@ 0896-6273.

Cinesi, C., L. Aeschbach, B. Yang and V. Dion (2016). "Contracting CAG/CTG repeats using the CRISPR-Cas9 nickase." <u>Nature Communications</u> **7**: 13272.

Clark, E., C. Strawser, K. Schadt, D. R. J. A. o. C. Lynch and T. Neurology (2019). "Identification of a novel missense mutation in Friedreich's ataxia–FXNW 168R."

Cleary, J. D., K. Nichol, Y. H. Wang and C. E. Pearson (2002). "Evidence of cis-acting factors in replication-mediated trinucleotide repeat instability in primate cells." <u>Nat Genet</u> **31**(1): 37-46.

Cleary, J. D. and C. E. Pearson (2005). "Replication fork dynamics and dynamic mutations: the fork-shift model of repeat instability." <u>Trends Genet</u> **21**(5): 272-280.

Cleary, J. D., S. Tome, A. Lopez Castel, G. B. Panigrahi, L. Foiry, K. A. Hagerman, H. Sroka, D. Chitayat, G. Gourdon and C. E. Pearson (2010). "Tissue- and age-specific DNA replication patterns at the CTG/CAG-expanded human myotonic dystrophy type 1 locus." <u>Nat Struct Mol Biol</u> **17**(9): 1079-1087.

Colak, D., N. Zaninovic, M. S. Cohen, Z. Rosenwaks, W.-Y. Yang, J. Gerhardt, M. D. Disney and S. R. J. S. Jaffrey (2014). "Promoter-bound trinucleotide repeat mRNA drives epigenetic silencing in fragile X syndrome." **343**(6174): 1002-1005.

Conrad, D. F., D. Pinto, R. Redon, L. Feuk, O. Gokcumen, Y. Zhang, J. Aerts, T. D. Andrews, C. Barnes and P. J. N. Campbell (2010). "Origins and functional impact of copy number variation in the human genome." J Nature **464**(7289): 704.

Consortium, I. H. G. S. (2001). "Initial sequencing and analysis of the human genome." <u>Nature</u> **409**(6822): 860.

Cossée, M., M. Schmitt, V. Campuzano, L. Reutenauer, C. Moutou, J.-L. Mandel and M. Koenig (1997). "Evolution of the Friedreich's ataxia trinucleotide repeat expansion: Founder effect and premutations." <u>Proceedings of the National Academy of Sciences</u> **94**(14): 7452.

Costa, I. P. D., B. C. Almeida, J. Sequeiros, A. Amorim and S. Martins (2019). "A Pipeline to Assess Disease-Associated Haplotypes in Repeat Expansion Disorders: The Example of MJD/SCA3 Locus." <u>Front Genet</u> **10**: 38.

Crespan, E., T. Czabany, G. Maga and U. Hubscher (2012). "Microhomology-mediated DNA strand annealing and elongation by human DNA polymerases lambda and beta on normal and repetitive DNA sequences." <u>Nucleic Acids Res</u> **40**(12): 5577-5590.

Crossley, M. P., M. Bocek and K. A. Cimprich (2019). "R-Loops as Cellular Regulators and Genomic Threats." <u>Mol Cell</u> **73**(3): 398-411.

Cumming, S. A., M. J. Hamilton, Y. Robb, H. Gregory, C. McWilliam, A. Cooper, B. Adam, J. McGhie, G. Hamilton, P. Herzyk, M. R. Tschannen, E. Worthey, R. Petty, B. Ballantyne, C. Scottish Myotonic Dystrophy, J. Warner, M. E. Farrugia, C. Longman and D. G. Monckton (2018). "De novo repeat interruptions are associated with reduced somatic instability and mild or absent clinical features in myotonic dystrophy type 1." <u>European journal of human genetics</u> : <u>EJHG</u> **26**(11): 1635-1647.

Cumming, S. A., M. J. Hamilton, Y. Robb, H. Gregory, C. McWilliam, A. Cooper, B. Adam, J. McGhie, G. Hamilton, P. Herzyk, M. R. Tschannen, E. Worthey, R. Petty, B. Ballantyne, J. Warner, M. E. Farrugia, C. Longman, D. G. Monckton and C. The Scottish Myotonic Dystrophy (2018). "De novo repeat interruptions are associated with reduced somatic instability and mild or absent clinical features in myotonic dystrophy type 1." <u>European Journal of Human Genetics</u> **26**(11): 1635-1647.

Cumming, S. A., C. Jimenez-Moreno, K. Okkersen, S. Wenninger, F. Daidj, F. Hogarth, R. Littleford, G. Gorman, G. Bassez, B. Schoser, H. Lochmuller, B. G. M. van Engelen and D. G. Monckton (2019). "Genetic determinants of disease severity in the myotonic dystrophy type 1 OPTIMISTIC cohort." <u>Neurology</u>.

Cummings, C. J., E. Reinstein, Y. Sun, B. Antalffy, Y.-h. Jiang, A. Ciechanover, H. T. Orr, A. L. Beaudet and H. Y. Zoghbi (1999). "Mutation of the E6-AP ubiquitin ligase reduces nuclear inclusion frequency while accelerating polyglutamine-induced pathology in SCA1 mice." <u>Neuron</u> **24**(4): 879-892 %@ 0896-6273.

Dandelot, E. and G. Gourdon (2018). "The flash-small-pool PCR: how to transform blotting and numerous hybridization steps into a simple denatured PCR." <u>Biotechniques</u> **64**(6): 262-265.

Dandelot, E. and S. Tomé (2017). "Genetic Modifiers of CAG. CTG Repeat Instability in Huntington's Disease Mouse Models." <u>Huntington's Disease: Molecular Pathogenesis and Current Models</u>: 1.

Dansithong, W., S. P. Jog, S. Paul, R. Mohammadzadeh, S. Tring, Y. Kwok, R. C. Fry, P. Marjoram, L. Comai and S. Reddy (2011). "RNA steady-state defects in myotonic dystrophy are linked to nuclear exclusion of SHARP." <u>EMBO reports</u> **12**(7): 735-742.

Dastidar, S., S. Ardui, K. Singh, D. Majumdar, N. Nair, Y. Fu, D. Reyon, E. Samara, M. F M. Gerli, A. F. Klein, W. De Schrijver, J. Tipanee, S. Seneca, W. Tulalamba, H. Wang, Yoke C. Chai, P. In't Veld, D. Furling, Francesco S. Tedesco, J. R. Vermeesch, J. K. Joung, M. K. Chuah and T. VandenDriessche (2018). "Efficient CRISPR/Cas9-mediated editing of trinucleotide repeat expansion in myotonic dystrophy patient-derived iPS and myogenic cells." <u>Nucleic Acids Research</u> **46**(16): 8275-8298.

Daughters, R. S., D. L. Tuttle, W. Gao, Y. Ikeda, M. L. Moseley, T. J. Ebner, M. S. Swanson and L. P. J. P. g. Ranum (2009). "RNA gain-of-function in spinocerebellar ataxia type 8." **5**(8): e1000600.

Day, J. W. and L. P. Ranum (2005). "RNA pathogenesis of the myotonic dystrophies." <u>Neuromuscular disorders</u> **15**(1): 5-16.

De Biase, I., A. Rasmussen, D. Endres, S. Al-Mahdawi, A. Monticelli, S. Cocozza, M. Pook and S. I. J. A. o. n. Bidichandani (2007). "Progressive GAA expansions in dorsal root ganglia of Friedreich's ataxia patients." **61**(1): 55-60.

de Koning, A. J., W. Gu, T. A. Castoe, M. A. Batzer and D. D. J. P. g. Pollock (2011). "Repetitive elements may comprise over two-thirds of the human genome." <u>PLoS genetics</u> **7**(12): e1002384.

De Michele, G., F. Cavalcanti, C. Criscuolo, L. Pianese, A. Monticelli, A. Filla and S. Cocozza (1998). "Parental gender, age at birth and expansion length influence GAA repeat intergenerational instability in the X25 gene: pedigree studies and analysis of sperm from patients with Friedreich's Ataxia." <u>Human Molecular Genetics</u> **7**(12): 1901-1906.

De Michele, G., F. Cavalcanti, C. Criscuolo, L. Pianese, A. Monticelli, A. Filla and S. J. H. m. g. Cocozza (1998). "Parental gender, age at birth and expansion length influence GAA repeat intergenerational instability in the X25 gene: pedigree studies and analysis of sperm from patients with Friedreich's ataxia." **7**(12): 1901-1906.

De Temmerman, N., K. Sermon, S. Seneca, M. De Rycke, P. Hilven, W. Lissens, A. Van Steirteghem and I. Liebaers (2004). "Intergenerational instability of the expanded CTG repeat in the DMPK gene: studies in human gametes and preimplantation embryos." <u>American journal of human genetics</u> **75**(2): 325-329.

Dean, N. L., S. L. Tan and A. Ao (2006). "Instability in the transmission of the myotonic dystrophy CTG repeat in human oocytes and preimplantation embryos." <u>Fertility and sterility</u> **86**(1): 98-105.

Devys, D., V. Biancalana, F. Rousseau, J. Boue, J. L. Mandel and I. Oberle (1992). "Analysis of full fragile X mutations in fetal tissues and monozygotic twins indicate that abnormal methylation and somatic heterogeneity are established early in development." <u>Am J Med Genet</u> **43**(1-2): 208-216.

Dewar, J. M. and J. C. Walter (2017). "Mechanisms of DNA replication termination." <u>Nature</u> reviews. Molecular cell biology **18**(8): 507-516.

Dion, V. (2014). "Tissue specificity in DNA repair: lessons from trinucleotide repeat instability." <u>Trends Genet</u> **30**(6): 220-229.

Dion, V., Y. Lin, L. Hubert, Jr., R. A. Waterland and J. H. Wilson (2008). "Dnmt1 deficiency promotes CAG repeat expansion in the mouse germline." <u>Human Molecular Genetics</u> **17**(9): 1306-1317.

Dion, V. and J. H. Wilson (2009). "Instability and chromatin structure of expanded trinucleotide repeats." <u>Trends in genetics : TIG</u> **25**(7): 288-297.

Ditch, S., M. C. Sammarco, A. Banerjee and E. Grabczyk (2009). "Progressive GAA.TTC repeat expansion in human cell lines." <u>PLoS genetics</u> **5**(10): e1000704-e1000704.

Dixon, M. J. and R. S. Lahue (2004). "DNA elements important for CAG· CTG repeat thresholds in Saccharomyces cerevisiae." <u>Nucleic acids research</u> **32**(4): 1289-1297.

Dragileva, E., A. Hendricks, A. Teed, T. Gillis, E. T. Lopez, E. C. Friedberg, R. Kucherlapati, W. Edelmann, K. L. Lunetta and M. E. MacDonald (2009). "Intergenerational and striatal CAG repeat instability in Huntington's disease knock-in mice involve different DNA repair genes." <u>Neurobiology of disease</u> **33**(1): 37-47.

Dryland, P. A., E. Doherty, J. M. Love and D. R. Love (2013). "Simple Repeat-Primed PCR Analysis of the Myotonic Dystrophy Type 1 Gene in a Clinical Diagnostics Environment." Journal of neurodegenerative diseases **2013**: 857564-857564.

Du, J., E. Campau, E. Soragni, C. Jespersen and J. M. Gottesfeld (2013). "Length-dependent CTG· CAG triplet-repeat expansion in myotonic dystrophy patient-derived induced pluripotent stem cells." <u>Human molecular genetics</u> **22**(25): 5276-5287.

Du, J., E. Campau, E. Soragni, C. Jespersen and J. M. Gottesfeld (2013). "Length-dependent CTG·CAG triplet-repeat expansion in myotonic dystrophy patient-derived induced pluripotent stem cells." <u>Human molecular genetics</u> **22**(25): 5276-5287.

Du, J., E. Campau, E. Soragni, S. Ku, J. W. Puckett, P. B. Dervan and J. M. Gottesfeld (2012). "Role of mismatch repair enzymes in GAA· TTC triplet-repeat expansion in Friedreich ataxia induced pluripotent stem cells." Journal of Biological Chemistry **287**(35): 29861-29872.

Dürr, A., M. Cossee, Y. Agid, V. Campuzano, C. Mignard, C. Penet, J.-L. Mandel, A. Brice and M. J. N. E. J. o. M. Koenig (1996). "Clinical and genetic abnormalities in patients with Friedreich's ataxia." **335**(16): 1169-1175.

E.Neville, C., M. S.Mahadeva, J. M.Barceló and R. Korneluk (1994). "High resolution genetic analysis suggests one ancestral predisposing haplotype for the origin of the myotonic dystrophy mutation." <u>Human Molecular Genetics</u> **3**(1): 45-51.

Eichler, E. E., J. J. Holden, B. W. Popovich, A. L. Reiss, K. Snow, S. N. Thibodeau, C. S. Richards, P. A. Ward and D. L. Nelson (1994). "Length of uninterrupted CGG repeats determines instability in the FMR1 gene." <u>Nature genetics</u> **8**(1): 88.

Emparanza, J. I., A. Lopez de Munain, M. H. Greene, A. Matheu, R. Fernandez-Torron and S. M. Gadalla (2018). "Cancer phenotype in myotonic dystrophy patients: Results from a metaanalysis." <u>Muscle Nerve</u> **58**(4): 517-522.

Ennis, S., A. Murray, G. Brightwell, N. E. Morton and P. A. Jacobs (2007). "Closely linked cisacting modifier of expansion of the CGG repeat in high risk FMR1 haplotypes." <u>Human</u> <u>mutation</u> **28**(12): 1216-1224.

Entezam, A., A. R. Lokanga, W. Le, G. Hoffman and K. Usdin (2010). "Potassium bromate, a potent DNA oxidizing agent, exacerbates germline repeat expansion in a fragile X premutation mouse model." <u>Hum Mutat</u> **31**(5): 611-616.

Evans-Galea, M. V., A. J. Hannan, N. Carrodus, M. B. Delatycki and R. Saffery (2013). "Epigenetic modifications in trinucleotide repeat diseases." <u>Trends in Molecular Medicine</u> **19**(11): 655-663.

Ezzatizadeh, V., R. M. Pinto, C. Sandi, M. Sandi, S. Al-Mahdawi, H. te Riele and M. A. Pook (2012). "The mismatch repair system protects against intergenerational GAA repeat instability in a Friedreich ataxia mouse model." <u>Neurobiology of disease</u> **46**(1): 165-171.

Fardaei, M., M. T. Rogers, H. M. Thorpe, K. Larkin, M. G. Hamshere, P. S. Harper and J. D. Brook (2002). "Three proteins, MBNL, MBLL and MBXL, co-localize in vivo with nuclear foci of expanded-repeat transcripts in DM1 and DM2 cells." <u>Human molecular genetics</u> **11**(7): 805-814.

Farrell, B. T. and R. S. Lahue (2006). "CAG CTG repeat instability in cultured human astrocytes." <u>Nucleic Acids Research</u> **34**(16): 4495-4505.

Feng, Y., D. Absher, D. E. Eberhart, V. Brown, H. E. Malter and S. T. Warren (1997). "FMRP associates with polyribosomes as an mRNP, and the I304N mutation of severe fragile X syndrome abolishes this association." <u>Mol Cell</u> **1**(1): 109-118.

Filippova, G. N. (2008). "Genetics and epigenetics of the multifunctional protein CTCF." <u>Curr</u> <u>Top Dev Biol</u> **80**: 337-360.

Filippova, G. N., C. P. Thienes, B. H. Penn, D. H. Cho, Y. J. Hu, J. M. Moore, T. R. Klesert, V. V. Lobanenkov and S. J. Tapscott (2001). "CTCF-binding sites flank CTG/CAG repeats and form a methylation-sensitive insulator at the DM1 locus." <u>Nat Genet</u> **28**(4): 335-343.

Flower, M., V. Lomeikaite, M. Ciosi, S. Cumming, F. Morales, K. Lo, D. Hensman Moss, L. Jones, P. Holmans, D. G. Monckton, S. J. Tabrizi, T.-H. Investigators and O. Consortium (2019). "MSH3 modifies somatic instability and disease severity in Huntington's and myotonic dystrophy type 1." <u>Brain</u> **142**(7): 1876-1886.

Foiry, L., L. Dong, C. Savouret, L. Hubert, H. Te Riele, C. Junien and G. Gourdon (2006). "Msh3 is a limiting factor in the formation of intergenerational CTG expansions in DM1 transgenic mice." <u>Human genetics</u> **119**(5): 520-526.

Follonier, C., J. Oehler, R. Herrador and M. Lopes (2013). "Friedreich's ataxia-associated GAA repeats induce replication-fork reversal and unusual molecular junctions." <u>Nat Struct Mol Biol</u> **20**(4): 486-494.

Freudenreich, C. H. (2018). "R-loops: targets for nuclease cleavage and repeat instability." <u>Current Genetics</u> **64**(4): 789-794.

Freudenreich, C. H., S. M. Kantrow and V. A. Zakian (1998). "Expansion and length-dependent fragility of CTG repeats in yeast." <u>Science</u> **279**(5352): 853-856.

Freyermuth, F., F. Rau, Y. Kokunai, T. Linke, C. Sellier, M. Nakamori, Y. Kino, L. Arandel, A. Jollet and C. Thibault (2016). "Splicing misregulation of SCN5A contributes to cardiac-conduction delay and heart arrhythmia in myotonic dystrophy." <u>Nature communications</u> **7**: 11067.

Frizzell, A., Jennifer H. G. Nguyen, Mark I. R. Petalcorin, Katherine D. Turner, Simon J. Boulton, Catherine H. Freudenreich and Robert S. Lahue (2014). "RTEL1 Inhibits Trinucleotide Repeat Expansions and Fragility." <u>Cell Reports</u> **6**(5): 827-835.

Fu, Y., A. Pizzuti, R. G. Fenwick, J. King, S. Rajnarayan, P. Dunne, J. Dubel, G. A. Nasser, T. Ashizawa and P. J. S. De Jong (1992). "An unstable triplet repeat in a gene related to myotonic muscular dystrophy." <u>Science</u> **255**(5049): 1256-1258.

Fu, Y.-H., D. P. Kuhl, A. Pizzuti, M. Pieretti, J. S. Sutcliffe, S. Richards, A. J. Verkert, J. J. Holden, R. G. Fenwick Jr and S. T. J. C. Warren (1991). "Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox." **67**(6): 1047-1058.

Fugier, C., A. F. Klein, C. Hammer, S. Vassilopoulos, Y. Ivarsson, A. Toussaint, V. Tosch, A. Vignaud, A. Ferry, N. Messaddeq, Y. Kokunai, R. Tsuburaya, P. de la Grange, D. Dembele, V. Francois, G. Precigout, C. Boulade-Ladame, M. C. Hummel, A. Lopez de Munain, N. Sergeant, A. Laquerriere, C. Thibault, F. Deryckere, D. Auboeuf, L. Garcia, P. Zimmermann, B. Udd, B. Schoser, M. P. Takahashi, I. Nishino, G. Bassez, J. Laporte, D. Furling and N. Charlet-Berguerand (2011). "Misregulated alternative splicing of BIN1 is associated with T tubule alterations and muscle weakness in myotonic dystrophy." <u>Nat Med</u> **17**(6): 720-725.

Fujigasaki, H., I. C. Verma, A. Camuzat, R. L. Margolis, C. Zander, A. S. Lebre, L. Jamot, R. Saxena, I. Anand, S. E. J. A. o. N. O. J. o. t. A. N. A. Holmes and t. C. N. Society (2001). "SCA12 is a rare locus for autosomal dominant cerebellar ataxia: a study of an Indian family." **49**(1): 117-121.

Gacy, A. M., G. Goellner, N. Juranić, S. Macura and C. T. McMurray (1995). "Trinucleotide repeats that expand in human disease form hairpin structures in vitro." <u>Cell</u> **81**(4): 533-540.

Gacy, A. M. and C. T. McMurray (1998). "Influence of hairpins on template reannealing at trinucleotide repeat duplexes: a model for slipped DNA." <u>Biochemistry</u> **37**(26): 9426-9434.

Gatchel, J. R. and H. Y. Zoghbi (2005). "Diseases of unstable repeat expansion: mechanisms and common principles." <u>Nature Reviews Genetics</u> **6**(10): 743 %@ 1471-0064.

Gecz, J., A. K. Gedeon, G. R. Sutherland and J. C. Mulley (1996). "Identification of the gene FMR2, associated with FRAXE mental retardation." <u>Nature Genetics</u> **13**(1): 105-108.

Gécz, J., B. A. Oostra, A. Hockey, P. Carbonell, G. Turner, E. A. Haan, G. R. Sutherland and J. C. Mulley (1997). "FMR2 Expression in Families with Fraxe Mental Retardation." <u>Human</u> <u>Molecular Genetics</u> **6**(3): 435-441.

Gécz, J. J. J. o. m. g. (2000). "FMR3 is a novel gene associated with FRAXE CpG island and transcriptionally silent in FRAXE full mutations." **37**(10): 782-784.

Genetic Modifiers of Huntington's Disease, C. (2015). "Identification of Genetic Factors that Modify Clinical Onset of Huntington's Disease." <u>Cell</u> **162**(3): 516-526.

Gerhardt, J., A. D. Bhalla, J. S. Butler, J. W. Puckett, P. B. Dervan, Z. Rosenwaks and M. Napierala (2016). "Stalled DNA replication forks at the endogenous GAA repeats drive repeat expansion in Friedreich's ataxia cells." <u>Cell reports</u> **16**(5): 1218-1227.

Gerhardt, J., M. J. Tomishima, N. Zaninovic, D. Colak, Z. Yan, Q. Zhan, Z. Rosenwaks, S. R. Jaffrey and C. L. Schildkraut (2014). "The DNA replication program is altered at the FMR1 locus in fragile X embryonic stem cells." <u>Mol Cell</u> **53**(1): 19-31.

Gerhardt, J., N. Zaninovic, Q. Zhan, A. Madireddy, S. L. Nolin, N. Ersalesi, Z. Yan, Z. Rosenwaks and C. L. Schildkraut (2014). "Cis-acting DNA sequence at a replication origin promotes repeat expansion to fragile X full mutation." <u>J Cell Biol</u> **206**(5): 599-607.

Ginno, P. A., P. L. Lott, H. C. Christensen, I. Korf and F. Chedin (2012). "R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters." <u>Mol Cell</u> **45**(6): 814-825.

Gold, M., K. Freon, A. Su, S. Lambert and C. Freudenreich (2017). "The Effects of Replication Fork Restart on CAG Repeat Instability in S. pombe." <u>The FASEB Journal</u> **31**(1_supplement): 591.595-591.595.

Goldberg, Y. P., C. T. McMurray, J. Zeisler, E. Almqvist, D. Sillence, F. Richards, A. M. Gacy, J. Buchanan, H. Telenius and M. R. Hayden (1995). "Increased instability of intermediate alleles in families with sporadic Huntington disease compared to similar sized intermediate alleles in the general population." <u>Hum Mol Genet</u> **4**(10): 1911-1918.

Gomes-Pereira, M., S. I. Bidichandani and D. G. Monckton (2004). Analysis of unstable triplet repeats using small-pool polymerase chain reaction. <u>Trinucleotide repeat protocols</u>, Springer: 61-76.

Gomes-Pereira, M., T. A. Cooper and G. Gourdon (2011). "Myotonic dystrophy mouse models: towards rational therapy development." <u>Trends in molecular medicine</u> **17**(9): 506-517.

Gomes-Pereira, M., M. T. Fortune, L. Ingram, J. P. McAbney and D. G. Monckton (2004). "Pms2 is a genetic enhancer of trinucleotide CAG·CTG repeat somatic mosaicism: implications for the mechanism of triplet repeat expansion." <u>Human Molecular Genetics</u> **13**(16): 1815-1825.

Gomes-Pereira, M., J. D. Hilley, F. Morales, B. Adam, H. E. James and D. G. Monckton (2014). "Disease-associated CAG·CTG triplet repeats expand rapidly in non-dividing mouse cells, but cell cycle arrest is insufficient to drive expansion." <u>Nucleic Acids Research</u> **42**(11): 7047-7056.

Gonitel, R., H. Moffitt, K. Sathasivam, B. Woodman, P. J. Detloff, R. L. M. Faull and G. P. Bates (2008). "DNA instability in postmitotic neurons." <u>Proceedings of the National Academy of Sciences of the United States of America</u> **105**(9): 3467-3472.

Goodman, F., S. Mundlos, Y. Muragaki, D. Donnai, M. Giovannucci-Uzielli, E. Lapi, F. Majewski, J. McGaughran, C. McKeown and W. J. P. o. t. N. A. o. S. Reardon (1997). "Synpolydactyly phenotypes correlate with size of expansions in HOXD13 polyalanine tract." **94**(14): 7458-7463.

Gottipati, P., T. N. Cassel, L. Savolainen and T. Helleday (2008). "Transcription-associated recombination is dependent on replication in Mammalian cells." <u>Molecular and cellular biology</u> **28**(1): 154-164.

Goula, A.-V., B. R. Berquist, D. M. Wilson III, V. C. Wheeler, Y. Trottier and K. J. P. g. Merienne (2009). "Stoichiometry of base excision repair proteins correlates with increased somatic CAG instability in striatum over cerebellum in Huntington's disease transgenic mice." **5**(12): e1000749.

Goula, A.-V., C. E. Pearson, J. Della Maria, Y. Trottier, A. E. Tomkinson, D. M. Wilson, 3rd and K. Merienne (2012). "The nucleotide sequence, DNA damage location, and protein stoichiometry influence the base excision repair outcome at CAG/CTG repeats." <u>Biochemistry</u> **51**(18): 3919-3932.

Goula, A. V. and K. Merienne (2013). "Abnormal base excision repair at trinucleotide repeats associated with diseases: a tissue-selective mechanism." <u>Genes (Basel)</u> **4**(3): 375-387.

Gourdon, G., F. Radvanyi, A. Lia, C. Duros, M. Blanche, M. Abitbol, C. Junien and H. Hofmann-Radvanyi (1997). "Moderate instability of a 55 CTG repeat in transgenic mice carrying a 45 kb genomic region from an affected DM patient." <u>Nat Genet</u> **15**: 190-192.

Gourdon, G., F. Radvanyi, A. S. Lia, C. Duros, M. Blanche, M. Abitbol, C. Junien and H. Hofmann-Radvanyi (1997). "Moderate intergenerational and somatic instability of a 55-CTG repeat in transgenic mice." <u>Nat Genet</u> **15**(2): 190-192.

Grabczyk, E. and K. Usdin (2000). "The GAA• TTC triplet repeat expanded in Friedreich's ataxia impedes transcription elongation by T7 RNA polymerase in a length and supercoil dependent manner." <u>Nucleic acids research</u> **28**(14): 2815-2822.

Greene, E., L. Mahishi, A. Entezam, D. Kumari and K. Usdin (2007). "Repeat-induced epigenetic changes in intron 1 of the frataxin gene and its consequences in Friedreich ataxia." <u>Nucleic acids research</u> **35**(10): 3383-3390.

Groenen, P. and B. Wieringa (1998). "Expanding complexity in myotonic dystrophy." <u>BioEssays</u> **20**(11): 901-912.

Groh, M. and N. Gromak (2014). "Out of balance: R-loops in human disease." <u>PLoS genetics</u> **10**(9): e1004630.

Groh, M., M. M. P. Lufino, R. Wade-Martins and N. Gromak (2014). "R-loops Associated with Triplet Repeat Expansions Promote Gene Silencing in Friedreich Ataxia and Fragile X Syndrome." <u>PLOS Genetics</u> **10**(5): e1004318.

Groh, M., M. M. P. Lufino, R. Wade-Martins and N. Gromak (2014). "R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome." <u>PLoS genetics</u> **10**(5): e1004318-e1004318.

Groh, W. J., M. R. Groh, C. Saha, J. C. Kincaid, Z. Simmons, E. Ciafaloni, R. Pourmand, R. F. Otten, D. Bhakta, G. V. Nair, M. M. Marashdeh, D. P. Zipes and R. M. Pascuzzi (2008). "Electrocardiographic abnormalities and sudden death in myotonic dystrophy type 1." <u>N Engl</u> <u>J Med</u> **358**(25): 2688-2697.

Gudde, A. E. E. G., S. J. van Heeringen, A. I. de Oude, I. D. G. van Kessel, J. Estabrook, E. T. Wang, B. Wieringa and D. G. Wansink (2017). "Antisense transcription of the myotonic dystrophy locus yields low-abundant RNAs with and without (CAG)n repeat." <u>RNA Biology</u> **14**(10): 1374-1388.

Guo, J., L. Gu, M. Leffak and G. M. Li (2016). "MutSbeta promotes trinucleotide repeat expansion by recruiting DNA polymerase beta to nascent (CAG)n or (CTG)n hairpins for error-prone DNA synthesis." <u>Cell Res</u> **26**(7): 775-786.

Hagerman, R. J., E. Berry-Kravis, H. C. Hazlett, D. B. Bailey Jr, H. Moine, R. F. Kooy, F. Tassone, I. Gantois, N. Sonenberg and J. L. J. N. r. D. p. Mandel (2017). "Fragile X syndrome." **3**: 17065.

Halabi, A., S. Ditch, J. Wang and E. Grabczyk (2012). "DNA mismatch repair complex MutSβ promotes GAA· TTC repeat expansion in human cells." <u>Journal of Biological Chemistry</u> **287**(35): 29958-29967.

Hamperl, S. and K. A. Cimprich (2016). "Conflict Resolution in the Genome: How Transcription and Replication Make It Work." <u>Cell</u> **167**(6): 1455-1467.

Hancock, J. M. J. B. (1996). "Simple sequences and the expanding genome." <u>Bioessays</u> **18**(5): 421-425.

Hancock, J. M. J. N. g. (1996). "Simple sequences in a 'minimal'genome." <u>Nature genetics</u> **14**(1): 14.

Hänsel-Hertsch, R., M. Di Antonio and S. Balasubramanian (2017). "DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential." <u>Nature Reviews Molecular</u> <u>Cell Biology</u> **18**: 279.

Harley, H. G., S. A. Rundle, J. C. MacMillan, J. Myring, J. D. Brook, S. Crow, W. Reardon, I. Fenton, D. J. Shaw and P. S. Harper (1993). "Size of the unstable CTG repeat sequence in relation to phenotype and parental transmission in myotonic dystrophy." <u>American journal of human genetics</u> **52**(6): 1164-1174.

Harmon, E. B., M. L. Harmon, T. D. Larsen, A. F. Paulson and M. B. Perryman (2008). "Myotonic dystrophy protein kinase is expressed in embryonic myocytes and is required for myotube formation." <u>Dev Dyn</u> **237**(9): 2353-2366.

Harper, P. S., H. G. Harley, W. Reardon and D. J. Shaw (1992). "Anticipation in myotonic dystrophy: new light on an old problem." <u>American journal of human genetics</u> **51**(1): 10-16.

He, Y., B. Vogelstein, V. E. Velculescu, N. Papadopoulos and K. W. Kinzler (2008). "The Antisense Transcriptomes of Human Cells." <u>Science</u> **322**(5909): 1855-1857.

Helderman-van den Enden, A. T., P. D. Maaswinkel-Mooij, E. Hoogendoorn, R. Willemsen, J. A. Maat-Kievit, M. Losekoot and B. A. Oostra (1999). "Monozygotic twin brothers with the fragile X syndrome: different CGG repeats and different mental capacities." J Med Genet **36**(3): 253-257.

Helmrich, A., M. Ballarino, E. Nudler and L. Tora (2013). "Transcription-replication encounters, consequences and genomic instability." <u>Nat Struct Mol Biol</u> **20**(4): 412-418.

Henricksen, L. A., S. Tom, Y. Liu and R. A. Bambara (2000). "Inhibition of flap endonuclease 1 by flap secondary structure and relevance to repeat sequence expansion." <u>Journal of Biological Chemistry</u> **275**(22): 16420-16427.

Hernández-Hernández, O., G. Sicot, D. M. Dinca, A. Huguet, A. Nicole, L. Buée, A. Munnich, N. Sergeant, G. Gourdon and M. Gomes-Pereira (2013). "Synaptic protein dysregulation in myotonic dystrophy type 1: Disease neuropathogenesis beyond missplicing." <u>Rare diseases</u> (Austin, Tex.) **1**: e25553-e25553.

Holmes, S. E., E. E. O'Hearn, M. G. McInnis, D. A. Gorelick-Feldman, J. J. Kleiderlein, C. Callahan, N. G. Kwak, R. G. Ingersoll-Ashworth, M. Sherr, A. J. Sumner, A. H. Sharp, U. Ananth, W. K. Seltzer, M. A. Boss, A.-M. Vieria-Saecker, J. T. Epplen, O. Riess, C. A. Ross

and R. L. Margolis (1999). "Expansion of a novel CAG trinucleotide repeat in the 5' region of PPP2R2B is associated with SCA12." <u>Nature Genetics</u> **23**(4): 391-392.

Hou, C., T. Zhang, L. Tian, J. Huang, L. Gu and G. M. Li (2011). "The Role of XPG in Processing (CAG)n/(CTG)n DNA Hairpins." <u>Cell Biosci</u> **1**(1): 11.

Hu, Y., Y. Hashimoto, T. Ishii, M. Rayle, K. Soga, N. Sato, M. Okita, M. Higashi, K. Ozaki and H. J. J. o. t. n. s. Mizusawa (2017). "Sequence configuration of spinocerebellar ataxia type 8 repeat expansions in a Japanese cohort of 797 ataxia subjects." **382**: 87-90.

Huguet, A., F. Medja, A. Nicole, A. Vignaud, C. Guiraud-Dogan, A. Ferry, V. Decostre, J.-Y. Hogrel, F. Metzger, A. Hoeflich, M. Baraibar, M. Gomes-Pereira, J. Puymirat, G. Bassez, D. Furling, A. Munnich and G. Gourdon (2012). "Molecular, physiological, and motor performance defects in DMSXL mice carrying >1,000 CTG repeats from the human DM1 locus." <u>PLoS genetics</u> **8**(11): e1003043-e1003043.

Hunter, A., C. Tsilfidis, G. Mettler, P. Jacob, M. Mahadevan, L. Surh and R. Korneluk (1992). "The correlation of age of onset with CTG trinucleotide repeat amplification in myotonic dystrophy." <u>J Med Genet</u> **29**(11): 774-779.

Ikeda, Y., J. C. Dalton, M. L. Moseley, K. L. Gardner, T. D. Bird, T. Ashizawa, W. K. Seltzer, M. Pandolfo, A. Milunsky and N. T. J. T. A. J. o. H. G. Potter (2004). "Spinocerebellar Ataxia Type 8: Molecular Genetic Comparisonsand Haplotype Analysis of 37 Families with Ataxia." **75**(1): 3-16.

Ikeda, Y., R. S. Daughters and L. P. W. Ranum (2008). "Bidirectional expression of the SCA8 expansion mutation: One mutation, two genes." <u>The Cerebellum</u> **7**(2): 150-158.

Ikeuchi, T., R. Koide, O. Onodera, H. Tanaka, M. Oyake, H. Takano and S. Tsuji (1995). "Dentatorubral-pallidoluysian atrophy (DRPLA). Molecular basis for wide clinical features of DRPLA." <u>Clinical neuroscience (New York, NY)</u> **3**(1): 23-27 %@ 1065-6766.

Illarioshkin, S. N., S. Igarashi, O. Onodera, E. D. Markova, N. N. Nikolskaya, H. Tanaka, T. Z. Chabrashwili, N. G. Insarova, K. Endo and I. A. Ivanova-Smolenskaya (1994). "Trinucleotide repeat length and rate of progression of Huntington's disease." <u>Annals of Neurology: Official</u> Journal of the American Neurological Association and the Child Neurology Society **36**(4): 630-635.

Imashimizu, M. and D. B. Lukatsky (2018). "Transcription pausing: biological significance of thermal fluctuations biased by repetitive genomic sequences." <u>Transcription</u> **9**(3): 196-203.

Imbert, G., C. Kretz, K. Johnson and J.-L. Mandel (1993). "Origin of the expansion mutation in myotonic dystrophy." <u>Nature Genetics</u> **4**(1): 72-76.

Ingram, M. A., H. T. Orr and H. B. Clark (2012). "Genetically engineered mouse models of the trinucleotide-repeat spinocerebellar ataxias." <u>Brain Res Bull</u> **88**(1): 33-42.

Ivančić-Baće, I., J. Al Howard and E. L. Bolt (2012). "Tuning in to Interference: R-Loops and Cascade Complexes in CRISPR Immunity." Journal of Molecular Biology **422**(5): 607-616.

Iwahashi, C., D. Yasui, H.-J. An, C. Greco, F. Tassone, K. Nannen, B. Babineau, C. B. Lebrilla, R. J. Hagerman and P. J. J. B. Hagerman (2005). "Protein composition of the intranuclear inclusions of FXTAS." **129**(1): 256-271.

lyer, R. R., A. Pluciennik, M. Napierala and R. D. J. A. r. o. b. Wells (2015). "DNA triplet repeat expansion and mismatch repair." <u>Annual review of biochemistry</u> **84**: 199-226.

Jansen, G., P. J. Groenen, D. Bächner, P. H. Jap, M. Coerwinkel, F. Oerlemans, W. van den Broek, B. Gohlsch, D. Pette and J. J. Plomp (1996). "Abnormal myotonic dystrophy protein kinase levels produce only mild myopathy in mice." <u>Nature genetics</u> **13**(3): 316.

Jansen, G., P. Willems, M. Coerwinkel, W. Nillesen, H. Smeets, L. Vits, C. Höweler, H. Brunner and B. Wieringa (1994). "Gonosomal mosaicism in myotonic dystrophy patients: involvement of mitotic events in (CTG) n repeat variation and selection against extreme expansion in sperm." <u>American journal of human genetics</u> **54**(4): 575.

Janzen, M., M. Moseley, K. Benzow, J. Day, M. Koob and L. Ranum (1999). <u>Limited expression</u> of SCA8 is consistent with cerebellar pathogenesis and toxic of function RNA model. AMERICAN JOURNAL OF HUMAN GENETICS, UNIV CHICAGO PRESS 5720 SOUTH WOODLAWN AVE, CHICAGO, IL 60637-1603 USA.

Jarem, D. A., L. V. Huckaby and S. Delaney (2010). "AGG interruptions in (CGG)(n) DNA repeat tracts modulate the structure and thermodynamics of non-B conformations in vitro." <u>Biochemistry</u> **49**(32): 6826-6837.

Jarem, D. A., N. R. Wilson and S. Delaney (2009). "Structure-Dependent DNA Damage and Repair in a Trinucleotide Repeat Sequence." <u>Biochemistry</u> **48**(28): 6655-6663.

Jeffreys, A. J., R. Neumann and V. J. C. Wilson (1990). "Repeat unit sequence variation in minisatellites: a novel source of DNA polymorphism for studying variation and mutation by single molecule analysis." <u>Cell</u> **60**(3): 473-485.

Jeffreys, A. J., V. Wilson and S. L. J. N. Thein (1985). "Hypervariable 'minisatellite'regions in human DNA." <u>Nature</u> **314**(6006): 67.

Jenjaroenpun, P., T. Wongsurawat, S. Sutheeworapong and V. A. Kuznetsov (2016). "R-loopDB: a database for R-loop forming sequences (RLFS) and R-loops." <u>Nucleic Acids</u> <u>Research</u> **45**(D1): D119-D127.

Jiang, H., A. Mankodi, M. S. Swanson, R. T. Moxley and C. A. Thornton (2004). "Myotonic dystrophy type 1 is associated with nuclear foci of mutant RNA, sequestration of muscleblind proteins and deregulated alternative splicing in neurons." <u>Hum Mol Genet</u> **13**(24): 3079-3088.

Jiricny, J. (2006). "The multifaceted mismatch-repair system." <u>Nat Rev Mol Cell Biol</u> **7**(5): 335-346.

Jones, K., C. Wei, P. lakova, E. Bugiardini, C. Schneider-Gold, G. Meola, J. Woodgett, J. Killian, N. A. Timchenko and L. T. Timchenko (2012). "GSK3β mediates muscle pathology in myotonic dystrophy." <u>The Journal of clinical investigation</u> **122**(12): 4461-4472.

Kaliman, P. and E. Llagostera (2008). "Myotonic dystrophy protein kinase (DMPK) and its role in the pathogenesis of myotonic dystrophy 1." <u>Cellular signalling</u> **20**(11): 1935-1941.

Keogh, N., K. Y. Chan, G.-M. Li and R. S. Lahue (2017). "MutSβ abundance and Msh3 ATP hydrolysis activity are important drivers of CTG•CAG repeat expansions." <u>Nucleic acids</u> research **45**(17): 10068-10078.

Kim, J. C., S. T. Harris, T. Dinter, K. A. Shah and S. M. Mirkin (2017). "The role of breakinduced replication in large-scale expansions of (CAG)(n)/(CTG)(n) repeats." <u>Nature Structural</u> <u>& Molecular Biology</u> **24**(1): 55-60.

Kim, N. and S. Jinks-Robertson (2012). "Transcription as a source of genome instability." <u>Nature Reviews Genetics</u> **13**: 204.

Kim, T.-M. and P. J. Park (2014). "A genome-wide view of microsatellite instability: old stories of cancer mutations revisited with new sequencing technologies." <u>Cancer research</u> **74**(22): 6377-6382.

Kimura, R., T. Morihara, T. Kudo, K. Kamino and M. J. N. I. Takeda (2011). "Association between CAG repeat length in the PPP2R2B gene and Alzheimer disease in the Japanese population." **487**(3): 354-357.

Kimura, T., M. Nakamori, J. D. Lueck, P. Pouliquin, F. Aoike, H. Fujimura, R. T. Dirksen, M. P. Takahashi, A. F. Dulhunty and S. Sakoda (2005). "Altered mRNA splicing of the skeletal muscle ryanodine receptor and sarcoplasmic/endoplasmic reticulum Ca2+-ATPase in myotonic dystrophy type 1." Hum Mol Genet **14**(15): 2189-2200.

Klesert, T. R., D. H. Cho, J. I. Clark, J. Maylie, J. Adelman, L. Snider, E. C. Yuen, P. Soriano and S. J. Tapscott (2000). "Mice deficient in Six5 develop cataracts: implications for myotonic dystrophy." <u>Nature genetics</u> **25**(1): 105.

Klug, W. S. (2009). Concepts of genetics. PEARSON.

Kogi, M., S. Fukushige, C. Lefevre, S. Hadano and J.-E. J. G. Ikeda (1997). "A novel tandem repeat sequence located on human chromosome 4p: isolation and characterization." <u>Genomics</u> **42**(2): 278-283.

Kononenko, A. V., T. Ebersole, K. M. Vasquez and S. M. Mirkin (2018). "Mechanisms of genetic instability caused by (CGG)(n) repeats in an experimental mammalian system." <u>Nature structural & molecular biology</u> **25**(8): 669-676.

Koob, M. D., M. L. Moseley, L. J. Schut, K. A. Benzow, T. D. Bird, J. W. Day and L. P. J. N. g. Ranum (1999). "An untranslated CTG expansion causes a novel form of spinocerebellar ataxia (SCA8)." **21**(4): 379.

Kotsantis, P., L. M. Silva, S. Irmscher, R. M. Jones, L. Folkes, N. Gromak and E. Petermann (2016). "Increased global transcription activity as a mechanism of replication stress in cancer." <u>Nature Communications</u> **7**: 13087.

Kovtun, I. V., K. O. Johnson and C. T. McMurray (2011). "Cockayne syndrome B protein antagonizes OGG1 in modulating CAG repeat length in vivo." <u>Aging (Albany NY)</u> **3**(5): 509-514.

Kovtun, I. V., Y. Liu, M. Bjoras, A. Klungland, S. H. Wilson and C. T. McMurray (2007). "OGG1 initiates age-dependent CAG trinucleotide expansion in somatic cells." <u>Nature</u> **447**(7143): 447-452.

Kozlowski, P., M. de Mezer and W. J. Krzyzosiak (2010). "Trinucleotide repeats in human genome and exome." <u>Nucleic acids research</u> **38**(12): 4027-4039.

Krasilnikova, M. M., M. L. Kireeva, V. Petrovic, N. Knijnikova, M. Kashlev and S. M. Mirkin (2007). "Effects of Friedreich's ataxia (GAA)n*(TTC)n repeats on RNA synthesis and stability." <u>Nucleic Acids Res</u> **35**(4): 1075-1084.

Krasilnikova, M. M. and S. M. Mirkin (2004). "Replication stalling at Friedreich's ataxia (GAA) n repeats in vivo." <u>Molecular and cellular biology</u> **24**(6): 2286-2295.

Kraus-Perrotta, C. and S. Lagalwar (2016). "Expansion, mosaicism and interruption: mechanisms of the CAG repeat mutation in spinocerebellar ataxia type 1." <u>Cerebellum & ataxias</u> **3**: 20-20.

Kremer, B., E. Almqvist, J. Theilmann, N. Spence, H. Telenius, Y. P. Goldberg and M. R. Hayden (1995). "Sex-dependent mechanisms for expansions and contractions of the CAG repeat on affected Huntington disease chromosomes." <u>American journal of human genetics</u> **57**(2): 343-350.

Ku, S., E. Soragni, E. Campau, E. A. Thomas, G. Altun, L. C. Laurent, J. F. Loring, M. Napierala and J. M. Gottesfeld (2010). "Friedreich's ataxia induced pluripotent stem cells model intergenerational GAATTC triplet repeat instability." <u>Cell Stem Cell</u> **7**(5): 631-637.

Kuemmerle, S., C. A. Gutekunst, A. M. Klein, X. J. Li, S. H. Li, M. F. Beal, S. M. Hersch and R. J. Ferrante (1999). "Huntingtin aggregates may not predict neuronal death in Huntington's disease." <u>Annals of neurology</u> **46**(6): 842-849 %@ 0364-5134.

Kumar, A., S. Agarwal and S. Pradhan (2015). "Haplotype analysis and LD detection at DM1 locus." <u>Gene</u> **567**(1): 45-50.

Kunst, C. B. and S. T. Warren (1994). "Cryptic and polar variation of the fragile X repeat could result in predisposing normal alleles." <u>Cell</u> **77**(6): 853-861.

Kuyumcu-Martinez, N. M., G.-S. Wang and T. A. Cooper (2007). "Increased steady-state levels of CUGBP1 in myotonic dystrophy 1 are due to PKC-mediated hyperphosphorylation." <u>Molecular cell</u> **28**(1): 68-78.

Lai, Y., H. Budworth, J. M. Beaver, N. L. S. Chan, Z. Zhang, C. T. McMurray and Y. Liu (2016). "Crosstalk between MSH2-MSH3 and polβ promotes trinucleotide repeat expansion during base excision repair." <u>Nature communications</u> **7**: 12465-12465.

Lai, Y., M. Xu, Z. Zhang and Y. Liu (2013). "Instability of CTG repeats is governed by the position of a DNA base lesion through base excision repair." <u>PloS one</u> **8**(2): e56960-e56960.

Landick, R. (2006). "The regulatory roles and mechanism of transcriptional pausing." <u>Biochem</u> <u>Soc Trans</u> **34**(Pt 6): 1062-1066.

Lang, W. H., J. E. Coats, J. Majka, G. L. Hura, Y. Lin, I. Rasnik and C. T. McMurray (2011). "Conformational trapping of mismatch recognition complex MSH2/MSH3 on repair-resistant DNA loops." <u>Proc Natl Acad Sci U S A</u> **108**(42): E837-844.

Latham, G. J., J. Coppinger, A. G. Hadd and S. L. Nolin (2014). "The role of AGG interruptions in fragile X repeat expansions: a twenty-year perspective." <u>Frontiers in genetics</u> **5**: 244-244.

Laurent, F.-X., A. Sureau, A. F. Klein, F. Trouslard, E. Gasnier, D. Furling and J. Marie (2011). "New function for the RNA helicase p68/DDX5 as a modifier of MBNL1 activity on expanded CUG repeats." <u>Nucleic Acids Research</u> **40**(7): 3159-3171.

Lavedan, C., H. Hofmann-Radvanyi, P. Shelbourne, J. P. Rabes, C. Duros, D. Savoy, I. Dehaupas, S. Luce, K. Johnson and C. Junien (1993). "Myotonic dystrophy: size- and sex-dependent dynamics of CTG meiotic instability, and somatic mosaicism." <u>Am J Hum Genet</u> **52**(5): 875-883.

Lee, B., K. Lee, S. Panda, R. Gonzales-Rojas, A. Chong, V. Bugay, H. M. Park, R. Brenner, N. Murthy and H. Y. Lee (2018). "Nanoparticle delivery of CRISPR into the brain rescues a mouse model of fragile X syndrome from exaggerated repetitive behaviours." <u>Nature Biomedical Engineering</u> **2**(7): 497-507.

Lee, J.-M., M. J. Chao, D. Harold, K. Abu Elneel, T. Gillis, P. Holmans, L. Jones, M. Orth, R. H. Myers, S. Kwak, V. C. Wheeler, M. E. MacDonald and J. F. Gusella (2017). "A modifier of Huntington's disease onset at the MLH1 locus." <u>Human molecular genetics</u> **26**(19): 3859-3867.

Lee, S. and M. S. Park (2002). "Human FEN-1 can process the 5'-flap DNA of CTG/CAG triplet repeat derived from human genetic diseases by length and sequence dependent manner." <u>Experimental & molecular medicine</u> **34**(4): 313.

Leffak, M. (2017). "Break-induced replication links microsatellite expansion to complex genome rearrangements." <u>Bioessays</u> **39**(8): 1700025.

Leonard, A. C. and M. Méchali (2013). "DNA replication origins." <u>Cold Spring Harbor</u> perspectives in biology **5**(10): a010116-a010116.

Li, L., H. Liu, P. Dong, D. Li, W. R. Legant, J. B. Grimm, L. D. Lavis, E. Betzig, R. Tjian and Z. Liu (2016). "Real-time imaging of Huntingtin aggregates diverting target search and gene transcription." <u>Elife</u> **5**: e17056 %@ 12050-17084X.

Li, Y. C., A. B. Korol, T. Fahima, A. Beiles and E. J. M. e. Nevo (2002). "Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review." <u>Molecular</u> <u>ecology</u> **11**(12): 2453-2465.

Lia, A.-S., H. Seznec, H. Hofmann-Radvanyi, F. Radvanyi, C. Duros, C. Saquet, M. Blanche, C. Junien and G. Gourdon (1998). "Somatic Instability of the CTG Repeat in Mice Transgenic for the Myotonic Dystrophy Region is Age Dependent But Not Correlated to the Relative Intertissue Transcription Levels and Proliferative Capacities." <u>Human Molecular Genetics</u> **7**(8): 1285-1291.

Libby, R. T., K. A. Hagerman, V. V. Pineda, R. Lau, D. H. Cho, S. L. Baccam, M. M. Axford, J. D. Cleary, J. M. Moore, B. L. Sopher, S. J. Tapscott, G. N. Filippova, C. E. Pearson and A. R. La Spada (2008). "CTCF cis-regulates trinucleotide repeat instability in an epigenetic manner: a novel basis for mutational hot spot determination." <u>PLoS Genet</u> **4**(11): e1000257.

Libby, R. T., D. G. Monckton, Y. H. Fu, R. A. Martinez, J. P. McAbney, R. Lau, D. D. Einum, K. Nichol, C. B. Ware, L. J. Ptacek, C. E. Pearson and A. R. La Spada (2003). "Genomic context drives SCA7 CAG repeat instability, while expressed SCA7 cDNAs are intergenerationally and somatically stable in transgenic mice." <u>Hum Mol Genet</u> **12**(1): 41-50.

Lieberman, A. P., V. G. Shakkottai and R. L. J. A. R. o. P. M. o. D. Albin (2019). "Polyglutamine repeats in neurodegenerative diseases." <u>Annual Review of Pathology: Mechanisms of Disease</u> **14**: 1-27.

Lin, C.-H., C.-M. Chen, Y.-T. Hou, Y.-R. Wu, H.-M. Hsieh-Li, M.-T. Su and G.-J. J. H. g. Lee-Chen (2010). "The CAG repeat in SCA12 functions as a cis element to up-regulate PPP2R2B expression." **128**(2): 205-212.

Lin, Y., S. Y. Dent, J. H. Wilson, R. D. Wells and M. Napierala (2010). "R loops stimulate genetic instability of CTG.CAG repeats." <u>Proc Natl Acad Sci U S A</u> **107**(2): 692-697.

Lin, Y., V. Dion and J. H. Wilson (2006). "Transcription promotes contraction of CAG repeat tracts in human cells." <u>Nature structural & molecular biology</u> **13**(2): 179.

Lin, Y., L. Hubert Jr and J. H. Wilson (2009). "Transcription destabilizes triplet repeats." <u>Molecular Carcinogenesis: Published in cooperation with the University of Texas MD Anderson</u> <u>Cancer Center</u> **48**(4): 350-361. Lin, Y., L. Hubert, Jr. and J. H. Wilson (2009). "Transcription destabilizes triplet repeats." <u>Mol</u> <u>Carcinog</u> **48**(4): 350-361.

Lin, Y., M. Leng, M. Wan and J. H. Wilson (2010). "Convergent transcription through a long CAG tract destabilizes repeats and induces apoptosis." <u>Molecular and cellular biology</u> **30**(18): 4435-4451.

Lin, Y. and J. H. Wilson (2007). "Transcription-induced CAG repeat contraction in human cells is mediated in part by transcription-coupled nucleotide excision repair." <u>Mol Cell Biol</u> **27**(17): 6209-6217.

Lin, Y. and J. H. Wilson (2011). "Transcription-induced DNA toxicity at trinucleotide repeats." <u>Cell Cycle</u> **10**(4): 611-618.

Lin, Y. and J. H. Wilson (2012). "Nucleotide Excision Repair, Mismatch Repair, and R-Loops Modulate Convergent Transcription-Induced Cell Death and Repeat Instability." <u>PLOS ONE</u> **7**(10): e46807.

Lipps, H. J. and D. Rhodes (2009). "G-quadruplex structures: in vivo evidence and function." <u>Trends in Cell Biology</u> **19**(8): 414-422.

Liu, G., X. Chen, J. J. Bissler, R. R. Sinden and M. Leffak (2010). "Replication-dependent instability at (CTG)•(CAG) repeat hairpins in human cells." <u>Nature Chemical Biology</u> **6**: 652.

Liu, G., X. Chen, Y. Gao, T. Lewis, J. Barthelemy and M. Leffak (2012). "Altered replication in human cells promotes DMPK (CTG)(n) \cdot (CAG)(n) repeat instability." <u>Molecular and cellular biology</u> **32**(9): 1618-1632.

Liu, K.-Y., Y.-C. Shyu, B. A. Barbaro, Y.-T. Lin, Y. Chern, L. M. Thompson, C.-K. James Shen and J. L. Marsh (2014). "Disruption of the nuclear membrane by perinuclear inclusions of mutant huntingtin causes cell-cycle re-entry and striatal cell death in mouse and cell models of Huntington's disease." <u>Human molecular genetics</u> **24**(6): 1602-1616 %@ 1460-2083.

Liu, Y., R. Prasad, W. A. Beard, E. W. Hou, J. K. Horton, C. T. McMurray and S. H. Wilson (2009). "Coordination between Polymerase β and FEN1 Can Modulate CAG Repeat Expansion." Journal of Biological Chemistry **284**(41): 28352-28366.

Llagostera, E. (2007). "Catalucci D Marti L Liesa M Camps M Ciaraldi TP Kondo R Reddy S Dillmann WH Palacin M Zorzano A Ruiz-Lozano P Gomis R Kaliman P. Role of myotonic dystrophy protein kinase (DMPK) in glucose homeostasis and muscle insulin action." <u>PLoS ONE</u> **2**: e1134.

Lo Scrudato, M., K. Poulard, C. Sourd, S. Tome, A. F. Klein, G. Corre, A. Huguet, D. Furling, G. Gourdon and A. Buj-Bello (2019). "Genome Editing of Expanded CTG Repeats within the Human DMPK Gene Reduces Nuclear RNA Foci in the Muscle of DM1 Mice." <u>Mol Ther</u>.

Logigian, E. L., W. B. Martens, R. T. t. Moxley, M. P. McDermott, N. Dilek, A. W. Wiegner, A. T. Pearson, C. A. Barbieri, C. L. Annis, C. A. Thornton and R. T. Moxley, 3rd (2010). "Mexiletine is an effective antimyotonia treatment in myotonic dystrophy type 1." <u>Neurology</u> **74**(18): 1441-1448.

Lokanga, R. A., A. Entezam, D. Kumari, D. Yudkin, M. Qin, C. B. Smith and K. Usdin (2013). "Somatic expansion in mouse and human carriers of fragile X premutation alleles." <u>Hum Mutat</u> **34**(1): 157-166. Lokanga, R. A., X.-N. Zhao and K. Usdin (2014). "The mismatch repair protein MSH2 is rate limiting for repeat expansion in a fragile X premutation mouse model." <u>Human mutation</u> **35**(1): 129-136.

Lone, W. G., S. Poornima, A. K. Meena, K. P. Rao and Q. J. J. o. M. N. Hasan (2014). "Role of dynamic and mitochondrial mutations in neurodegenerative diseases with ataxia: lower repeats and LNAs at multiple loci as alternative pathogenesis." **54**(4): 837-847.

Long, A., J. S. Napierala, U. Polak, L. Hauser, A. H. Koeppen, D. R. Lynch and M. Napierala (2017). "Somatic instability of the expanded GAA repeats in Friedreich's ataxia." <u>PloS one</u> **12**(12): e0189990-e0189990.

Loomis, E. W., L. A. Sanz, F. Chédin and P. J. Hagerman (2014). "Transcription-associated R-loop formation across the human FMR1 CGG-repeat region." <u>PLoS genetics</u> **10**(4): e1004294-e1004294.

López Castel, A., J. D. Cleary and C. E. Pearson (2010). "Repeat instability as the basis for human diseases and as a potential target for therapy." <u>Nature Reviews Molecular Cell Biology</u> **11**: 165.

Lopez-Morato, M., J. D. Brook and M. Wojciechowska (2018). "Small Molecules Which Improve Pathogenesis of Myotonic Dystrophy Type 1." <u>Front Neurol</u> **9**: 349.

Lu, W.-T., B. R. Hawley, G. L. Skalka, R. A. Baldock, E. M. Smith, A. S. Bader, M. Malewicz, F. Z. Watts, A. Wilczynska and M. Bushell (2018). "Drosha drives the formation of DNA:RNA hybrids around DNA break sites to facilitate DNA repair." <u>Nature communications</u> **9**(1): 532-532.

Lynch, H. T. (1999). "Hereditary nonpolyposis colorectal cancer (HNPCC)." <u>Cytogenetic and</u> <u>Genome Research</u> **86**(2): 130-135.

Machuca-Tzili, L., H. Thorpe, T. E. Robinson, C. Sewry and J. D. Brook (2006). "Flies deficient in Muscleblind protein model features of myotonic dystrophy with altered splice forms of Z-band associated transcripts." <u>Hum Genet</u> **120**(4): 487-499.

Mahadevan, M., C. Tsilfidis, L. Sabourin, G. Shutler, C. Amemiya, G. Jansen, C. Neville, M. Narang, J. Barceló and K. J. S. O'Hoy (1992). "Myotonic dystrophy mutation: an unstable CTG repeat in the 3'untranslated region of the gene." <u>Science</u> **255**(5049): 1253-1255.

Maia, N., J. R. Loureiro, B. Oliveira, I. Marques, R. Santos, P. Jorge and S. Martins (2017). "Contraction of fully expanded FMR1 alleles to the normal range: predisposing haplotype or rare events?" <u>J Hum Genet</u> **62**(2): 269-275.

Maizels, N. (2006). "Dynamic roles for G4 DNA in the biology of eukaryotic cells." <u>Nature</u> <u>structural & molecular biology</u> **13**(12): 1055.

Malter, H. E., J. C. Iber, R. Willemsen, E. de Graaff, J. C. Tarleton, J. Leisti, S. T. Warren and B. A. Oostra (1997). "Characterization of the full fragile X syndrome mutation in fetal gametes." <u>Nat Genet</u> **15**(2): 165-169.

Mankodi, A., E. Logigian, L. Callahan, C. McClain, R. White, D. Henderson, M. Krym and C. A. Thornton (2000). "Myotonic dystrophy in transgenic mice expressing an expanded CUG repeat." <u>Science</u> **289**(5485): 1769-1773.

Mankodi, A., M. P. Takahashi, H. Jiang, C. L. Beck, W. J. Bowers, R. T. Moxley, S. C. Cannon and C. A. Thornton (2002). "Expanded CUG repeats trigger aberrant splicing of CIC-1 chloride

channel pre-mRNA and hyperexcitability of skeletal muscle in myotonic dystrophy." <u>Mol Cell</u> **10**(1): 35-44.

Manley, K., T. L. Shirley, L. Flaherty and A. Messer (1999). "Msh2 deficiency prevents in vivo somatic instability of the CAG repeat in Huntington disease transgenic mice." <u>Nature genetics</u> **23**(4): 471.

Margolis, D. G. A., K. Amanda and L. Russell (2019). "Huntington Disease-Like 2."

Marians, K. J. (2018). "Lesion Bypass and the Reactivation of Stalled Replication Forks." <u>Annu</u> <u>Rev Biochem</u> **87**: 217-238.

Marteyn, A., Y. Maury, Morgane M. Gauthier, C. Lecuyer, R. Vernet, Jérôme A. Denis, G. Pietu, M. Peschanski and C. Martinat (2011). "Mutant Human Embryonic Stem Cells Reveal Neurite and Synapse Formation Defects in Type 1 Myotonic Dystrophy." <u>Cell Stem Cell</u> **8**(4): 434-444.

Martinez-Cerdeno, V., M. Lechpammer, P. J. Hagerman and R. J. M. d. o. j. o. t. M. D. S. Hagerman (2017). "Two FMR1 premutation cases without nuclear inclusions." **32**(9): 1328.

Martins, S., A. I. Seixas, P. Magalhães, P. Coutinho, J. Sequeiros and I. J. A. J. o. M. G. P. B. N. G. Silveira (2005). "Haplotype diversity and somatic instability in normal and expanded SCA8 alleles." **139**(1): 109-114.

Martorell, L., A. M. Cobo, M. Baiget, M. Naudo, J. J. Poza and J. Parra (2007). "Prenatal diagnosis in myotonic dystrophy type 1. Thirteen years of experience: implications for reproductive counselling in DM1 families." <u>Prenat Diagn</u> **27**(1): 68-72.

Martorell, L., J. Gámez, M. L. Cayuela, F. K. Gould, J. P. McAbney, T. Ashizawa, D. G. Monckton and M. Baiget (2004). "Germline mutational dynamics in myotonic dystrophy type 1 males." <u>Neurology</u> **62**(2): 269.

Martorell, L., K. Johnson, C. A. Boucher and M. Baiget (1997). "Somatic Instability of the Myotonic Dystrophy (CTG)n Repeat during Human Fetal Development." <u>Human Molecular</u> <u>Genetics</u> **6**(6): 877-880.

Martorell, L., D. G. Monckton, J. Gamez and M. Baiget (2000). "Complex patterns of male germline instability and somatic mosaicism in myotonic dystrophy type 1." <u>European Journal of Human Genetics</u> **8**(6): 423-430.

Masumoto, H., H. Masukata, Y. Muro, N. Nozaki and T. Okazaki (1989). "A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite." <u>The Journal of Cell Biology</u> **109**(5): 1963.

Mateo, I., J. Llorca, V. Volpini, J. Corral, J. Berciano and O. J. A. n. s. Combarros (2004). "Expanded GAA repeats and clinical variation in Friedreich's ataxia." **109**(1): 75-78.

Matera, I., T. Bachetti, F. Puppo, M. Di Duca, F. Morandi, G. Casiraghi, M. Cilio, R. Hennekam, R. Hofstra and J. J. J. o. m. g. Schöber (2004). "PHOX2B mutations and polyalanine expansions correlate with the severity of the respiratory phenotype and associated symptoms in both congenital and late onset Central Hypoventilation syndrome." <u>J Journal of medical genetics</u> **41**(5): 373-380.

Maya-Mendoza, A., P. Moudry, J. M. Merchut-Maya, M. Lee, R. Strauss and J. Bartek (2018). "High speed of fork progression induces DNA replication stress and genomic instability." <u>Nature</u> **559**(7713): 279-284. McGinty, R. J. and S. M. Mirkin (2018). "Cis- and Trans-Modifiers of Repeat Expansions: Blending Model Systems with Human Genetics." <u>Trends Genet</u> **34**(6): 448-465.

McIvor, E. I., U. Polak and M. Napierala (2010). "New insights into repeat instability: role of RNA•DNA hybrids." <u>RNA biology</u> **7**(5): 551-558.

Menon, R. P., S. Nethisinghe, S. Faggiano, T. Vannocci, H. Rezaei, S. Pemble, M. G. Sweeney, N. W. Wood, M. B. Davis and A. Pastore (2013). "The role of interruptions in polyQ in the pathology of SCA1." <u>PLoS genetics</u> **9**(7): e1003648.

Merrill, R. A., A. M. Slupe and S. Strack (2012). Spinocerebellar Ataxia Type 12 (SCA 12): Clinical Features and Pathogenetic Mechanisms. <u>Spinocerebellar Ataxia</u>, IntechOpen.

Meservy, J. L., R. G. Sargent, R. R. Iyer, F. Chan, G. J. McKenzie, R. D. Wells and J. H. Wilson (2003). "Long CTG tracts from the myotonic dystrophy gene induce deletions and rearrangements during recombination at the APRT locus in CHO cells." <u>Molecular and cellular biology</u> **23**(9): 3152-3162.

Messaed, C. and G. A. Rouleau (2009). "Molecular mechanisms underlying polyalanine diseases." <u>Neurobiology of Disease</u> **34**(3): 397-405.

Michel, L., A. Huguet-Lachon and G. Gourdon (2015). "Sense and antisense DMPK RNA foci accumulate in DM1 tissues during development." <u>PLoS One</u> **10**(9): e0137620.

Miller, J. W., C. R. Urbinati, P. Teng-Umnuay, M. G. Stenberg, B. J. Byrne, C. A. Thornton and M. S. Swanson (2000). "Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy." <u>Embo j</u> **19**(17): 4439-4448.

Mirkin, E. V. and S. M. Mirkin (2007). "Replication Fork Stalling at Natural Impediments." <u>Microbiology and Molecular Biology Reviews</u> **71**(1): 13-35.

Mirkin, S. M. J. N. (2007). "Expandable DNA repeats and human disease." 447(7147): 932.

Mitas, M., A. Yu, J. Dill and I. S. Haworth (1995). "The trinucleotide repeat sequence d(CGG)15 forms a heat-stable hairpin containing Gsyn. Ganti base pairs." <u>Biochemistry</u> **34**(39): 12803-12811.

Mittelman, D., C. Moye, J. Morton, K. Sykoudis, Y. Lin, D. Carroll and J. H. Wilson (2009). "Zinc-finger directed double-strand breaks within CAG repeat tracts promote repeat instability in human cells." <u>Proceedings of the National Academy of Sciences</u> **106**(24): 9607-9612.

Mollersen, L., A. D. Rowe, J. L. Illuzzi, G. A. Hildrestrand, K. J. Gerhold, L. Tveteras, A. Bjolgerud, D. M. Wilson, 3rd, M. Bjoras and A. Klungland (2012). "Neil1 is a genetic modifier of somatic and germline CAG trinucleotide repeat instability in R6/1 mice." <u>Hum Mol Genet</u> **21**(22): 4939-4947.

Monckton, D. G., L. J. Wong, T. Ashizawa and C. T. Caskey (1995). "Somatic mosaicism, germline expansions, germline reversions and intergenerational reductions in myotonic dystrophy males: small pool PCR analyses." <u>Hum Mol Genet</u> **4**(1): 1-8.

Mondello, C., L. Pirzio, C. M. Azzalin and E. Giulotto (2000). "Instability of Interstitial Telomeric Sequences in the Human Genome." <u>Genomics</u> **68**(2): 111-117.

Monteys, A. M., S. A. Ebanks, M. S. Keiser and B. L. Davidson (2017). "CRISPR/Cas9 Editing of the Mutant Huntingtin Allele In Vitro and In Vivo." <u>Mol Ther</u> **25**(1): 12-23.

Morales, F., J. M. Couto, C. F. Higham, G. Hogg, P. Cuenca, C. Braida, R. H. Wilson, B. Adam, G. Del Valle and R. Brian (2012). "Somatic instability of the expanded CTG triplet repeat in myotonic dystrophy type 1 is a heritable quantitative trait and modifier of disease severity." <u>Human molecular genetics</u> **21**(16): 3558-3567.

Morales, F., M. Vásquez, P. Cuenca, D. Campos, C. Santamaría, G. Del Valle, R. Brian, M. Sittenfeld and D. G. Monckton (2015). "Parental age effects, but no evidence for an intrauterine effect in the transmission of myotonic dystrophy type 1." <u>European Journal of Human Genetics</u> **23**(5): 646.

Morales, F., M. Vasquez, C. Santamaria, P. Cuenca, E. Corrales and D. G. Monckton (2016). "A polymorphism in the MSH3 mismatch repair gene is associated with the levels of somatic instability of the expanded CTG repeat in the blood DNA of myotonic dystrophy type 1 patients." <u>DNA Repair (Amst)</u> **40**: 57-66.

Moseley, M. L., L. J. Schut, T. D. Bird, M. D. Koob, J. W. Day and L. P. J. H. m. g. Ranum (2000). "SCA8 CTG repeat: en masse contractions in sperm and intergenerational sequence changes may play a role in reduced penetrance." **9**(14): 2125-2130.

Moseley, M. L., T. Zu, Y. Ikeda, W. Gao, A. K. Mosemiller, R. S. Daughters, G. Chen, M. R. Weatherspoon, H. B. Clark and T. J. J. N. g. Ebner (2006). "Bidirectional expression of CUG and CAG expansion transcripts and intranuclear polyglutamine inclusions in spinocerebellar ataxia type 8." **38**(7): 758.

Moutou, C., M.-C. Vincent, V. Biancalana and J.-L. Mandel (1997). "Transition from Premutation to Full Mutation in Fragile X Syndrome is Likely to be Prezygotic." <u>Human</u> <u>Molecular Genetics</u> **6**(7): 971-979.

Mulders, S. A., W. J. van den Broek, T. M. Wheeler, H. J. Croes, P. van Kuik-Romeijn, S. J. de Kimpe, D. Furling, G. J. Platenburg, G. Gourdon, C. A. Thornton, B. Wieringa and D. G. Wansink (2009). "Triplet-repeat oligonucleotide-mediated reversal of RNA toxicity in myotonic dystrophy." <u>Proc Natl Acad Sci U S A</u> **106**(33): 13915-13920.

Mulvihill, D. J., K. Nichol Edamura, K. A. Hagerman, C. E. Pearson and Y. H. Wang (2005). "Effect of CAT or AGG interruptions and CpG methylation on nucleosome assembly upon trinucleotide repeats on spinocerebellar ataxia, type 1 and fragile X syndrome." <u>J Biol Chem</u> **280**(6): 4498-4503.

Muragaki, Y., S. Mundlos, J. Upton and B. R. Olsen (1996). "Altered Growth and Branching Patterns in Synpolydactyly Caused by Mutations in HOXD13." <u>Science</u> **272**(5261): 548.

Musova, Z., R. Mazanec, A. Krepelova, E. Ehler, J. Vales, R. Jaklova, T. Prochazka, P. Koukal, T. Marikova, J. Kraus, M. Havlovicova and Z. Sedlacek (2009). "Highly unstable sequence interruptions of the CTG repeat in the myotonic dystrophy gene." <u>Am J Med Genet A</u> **149A**(7): 1365-1374.

Musova, Z., R. Mazanec, A. Krepelova, E. Ehler, J. Vales, R. Jaklova, T. Prochazka, P. Koukal, T. Marikova, J. Kraus, M. Havlovicova and Z. Sedlacek (2009). "Highly unstable sequence interruptions of the CTG repeat in the myotonic dystrophy gene." <u>American Journal of Medical Genetics Part A</u> **149A**(7): 1365-1374.

Nageshwaran, S. and R. Festenstein (2015). "Epigenetics and Triplet-Repeat Neurological Diseases." <u>Neurology</u> **6**(262).

Nakamori, M., T. Kimura, H. Fujimura, M. P. Takahashi and S. Sakoda (2007). "Altered mRNA splicing of dystrophin in type 1 myotonic dystrophy." <u>Muscle Nerve</u> **36**(2): 251-257.

Nakamori, M., T. Kimura, T. Kubota, T. Matsumura, H. Sumi, H. Fujimura, M. Takahashi and S. Sakoda (2008). "Aberrantly spliced α -dystrobrevin alters α -syntrophin binding in myotonic dystrophy type 1." <u>Neurology</u> **70**(9): 677-685.

Nakamori, M., C. E. Pearson and C. A. J. H. m. g. Thornton (2010). "Bidirectional transcription stimulates expansion and contraction of expanded (CTG)•(CAG) repeats." <u>Human molecular genetics</u> **20**(3): 580-588.

Nakamori, M., K. Sobczak, A. Puwanant, S. Welle, K. Eichinger, S. Pandya, J. Dekdebrun, C. R. Heatwole, M. P. McDermott, T. Chen, M. Cline, R. Tawil, R. J. Osborne, T. M. Wheeler, M. S. Swanson, R. T. Moxley, 3rd and C. A. Thornton (2013). "Splicing biomarkers of disease severity in myotonic dystrophy." <u>Annals of neurology</u> **74**(6): 862-872.

Nakamori, M., K. Taylor, H. Mochizuki, K. Sobczak and M. P. Takahashi (2016). "Oral administration of erythromycin decreases RNA toxicity in myotonic dystrophy." <u>Annals of Clinical and Translational Neurology</u> **3**(1): 42-54.

Nakamoto, M., S. Nakano, S. Kawashima, M. Ihara, Y. Nishimura, A. Shinde and A. Kakizuka (2002). "Unequal Crossing-over in Unique PABP2 Mutations in Japanese Patients: A Possible Cause of Oculopharyngeal Muscular Dystrophy." JAMA Neurology **59**(3): 474-477.

Nakatani, R., M. Nakamori, H. Fujimura, H. Mochizuki and M. P. Takahashi (2015). "Large expansion of CTG• CAG repeats is exacerbated by MutSβ in human cells." <u>Scientific reports</u> **5**: 11020.

Napierala, M., P. Parniewski, A. Pluciennik and R. D. Wells (2002). "Long CTG.CAG repeat sequences markedly stimulate intramolecular recombination." <u>J Biol Chem</u> **277**(37): 34087-34100.

Nedelcheva, M. N., A. Roguev, L. B. Dolapchiev, A. Shevchenko, H. B. Taskov, A. Shevchenko, A. F. Stewart and S. S. Stoynov (2005). "Uncoupling of unwinding from DNA synthesis implies regulation of MCM helicase by Tof1/Mrc1/Csm3 checkpoint complex." <u>J Mol Biol</u> **347**(3): 509-521.

Neil, A. J., M. U. Liang, A. N. Khristich, K. A. Shah and S. M. Mirkin (2018). "RNA–DNA hybrids promote the expansion of Friedreich's ataxia (GAA)n repeats via break-induced replication." <u>Nucleic Acids Research</u> **46**(7): 3487-3497 %@ 0305-1048.

Nestor, C. E. and D. G. Monckton (2011). "Correlation of inter-locus polyglutamine toxicity with CAG•CTG triplet repeat expandability and flanking genomic DNA GC content." <u>PloS one</u> **6**(12): e28260-e28260.

Nichol Edamura, K., M. R. Leonard and C. E. Pearson (2005). "Role of replication and CpG methylation in fragile X syndrome CGG deletions in primate cells." <u>American journal of human genetics</u> **76**(2): 302-311.

O'Hearn, E., S. Holmes, P. C. Calvert, C. A. Ross and R. L. J. N. Margolis (2001). "SCA-12: tremor with cerebellar and cortical atrophy is associated with a CAG repeat expansion." **56**(3): 299-303.

Oberle, I., F. Rousseau, D. Heitz, C. Kretz, D. Devys, A. Hanauer, J. Boue, M. F. Bertheas and J. Mandel (1991). "Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome." <u>Science</u>: 1097-1102.

Ohshima, K., L. Montermini, R. D. Wells and M. Pandolfo (1998). "Inhibitory effects of expanded GAA.TTC triplet repeats from intron I of the Friedreich ataxia gene on transcription and replication in vivo." <u>J Biol Chem</u> **273**(23): 14588-14595.

Okkersen, K., C. Jimenez-Moreno, S. Wenninger, F. Daidj, J. Glennon, S. Cumming, R. Littleford, D. G. Monckton, H. Lochmuller, M. Catt, C. G. Faber, A. Hapca, P. T. Donnan, G. Gorman, G. Bassez, B. Schoser, H. Knoop, S. Treweek and B. G. M. van Engelen (2018). "Cognitive behavioural therapy with optional graded exercise therapy in patients with severe fatigue with myotonic dystrophy type 1: a multicentre, single-blind, randomised trial." <u>Lancet Neurol</u> **17**(8): 671-680.

Orr, H. T. (2001). "Beyond the Qs in the polyglutamine diseases." <u>Genes & Development</u> **15**(8): 925-932 %@ 0890-9369.

Orr, H. T. and H. Y. Zoghbi (2007). "Trinucleotide repeat disorders." <u>Annu. Rev. Neurosci.</u> **30**: 575-621 %@ 0147-0006X.

Oussatcheva, E. A., V. I. Hashem, Y. Zou, R. R. Sinden and V. N. Potaman (2001). "Involvement of the nucleotide excision repair protein UvrA in instability of CAG*CTG repeat sequences in Escherichia coli." J Biol Chem **276**(33): 30878-30884.

Overby, S. J., E. Cerro-Herreros, B. Llamusi and R. Artero (2018). "RNA-mediated therapies in myotonic dystrophy." <u>Drug Discov Today</u> **23**(12): 2013-2022.

Overend, G., C. Legare, J. Mathieu, L. Bouchard, C. Gagnon and D. G. Monckton (2019). "Allele length of the DMPK CTG repeat is a predictor of progressive myotonic dystrophy type 1 phenotypes." <u>Hum Mol Genet</u> **28**(13): 2245-2254.

Owen, B. A., Z. Yang, M. Lai, M. Gajec, J. D. Badger, 2nd, J. J. Hayes, W. Edelmann, R. Kucherlapati, T. M. Wilson and C. T. McMurray (2005). "(CAG)(n)-hairpin DNA binds to Msh2-Msh3 and changes properties of mismatch recognition." <u>Nat Struct Mol Biol</u> **12**(8): 663-670.

P.Leeflang, E. and N. Arnhelm (1995). "A novel repeat structure at the myotonic dystrophy locus in a 37 repeat allele with unexpectedly high stability." <u>Human Molecular Genetics</u> **4**(1): 135-136.

Pace, J. K., 2nd and C. Feschotte (2007). "The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage." <u>Genome research</u> **17**(4): 422-432.

Pall, G. S., K. J. Johnson and G. L. Smith (2003). "Abnormal contractile activity and calcium cycling in cardiac myocytes isolated from DMPK knockout mice." <u>Physiological genomics</u> **13**(2): 139-146.

Pandolfo, M. J. J. o. n. (2009). "Friedreich ataxia: the clinical picture." **256**(1): 3-8.

Pantic, B., E. Trevisan, A. Citta, M. P. Rigobello, O. Marin, P. Bernardi, S. Salvatori and A. Rasola (2013). "Myotonic dystrophy protein kinase (DMPK) prevents ROS-induced cell death by assembling a hexokinase II-Src complex on the mitochondrial surface." <u>Cell death & disease</u> **4**(10): e858-e858.

Parniewski, P., A. Bacolla, A. Jaworski and R. D. Wells (1999). "Nucleotide excision repair affects the stability of long transcribed (CTG*CAG) tracts in an orientation-dependent manner in Escherichia coli." <u>Nucleic Acids Res</u> **27**(2): 616-623.
Parsons, M. A., R. R. Sinden and M. G. Izban (1998). "Transcriptional properties of RNA polymerase II within triplet repeat-containing DNA from the human myotonic dystrophy and fragile X loci." <u>J Biol Chem</u> **273**(41): 26998-27008.

Paul, S., W. Dansithong, D. Kim, J. Rossi, N. J. Webster, L. Comai and S. Reddy (2006). "Interaction of musleblind, CUG-BP1 and hnRNP H proteins in DM1-associated aberrant IR splicing." <u>The EMBO journal</u> **25**(18): 4271-4283.

Paulson, H. (2018). Repeat expansion diseases. <u>Handbook of clinical neurology</u>, Elsevier. **147:** 105-123.

Pearson, C. E., K. N. Edamura and J. D. Cleary (2005). "Repeat instability: mechanisms of dynamic mutations." <u>Nature Reviews Genetics</u> **6**(10): 729-742.

Pearson, C. E., K. N. Edamura and J. D. J. N. R. G. Cleary (2005). "Repeat instability: mechanisms of dynamic mutations." <u>Nature Reviews Genetics</u> **6**(10): 729.

Pearson, C. E., E. E. Eichler, D. Lorenzetti, S. F. Kramer, H. Y. Zoghbi, D. L. Nelson and R. R. Sinden (1998). "Interruptions in the Triplet Repeats of SCA1 and FRAXA Reduce the Propensity and Complexity of Slipped Strand DNA (S-DNA) Formation." <u>Biochemistry</u> **37**(8): 2701-2708.

Peier, A. M. and D. L. Nelson (2002). "Instability of a Premutation-Sized CGG Repeat in FMR1 YAC Transgenic Mice." <u>Genomics</u> **80**(4): 423-432.

Pelletier, R., M. M. Krasilnikova, G. M. Samadashwily, R. Lahue and S. M. Mirkin (2003). "Replication and expansion of trinucleotide repeats in yeast." <u>Molecular and cellular biology</u> **23**(4): 1349-1357.

Pešović, J., S. Perić, M. Brkušanin, G. Brajušković, V. Rakočević-Stojanović and D. Savić-Pavićević (2017). "Molecular genetic and clinical characterization of myotonic dystrophy type 1 patients carrying variant repeats within DMPK expansions." <u>neurogenetics</u> **18**(4): 207-218.

Pešović, J., S. Perić, M. Brkušanin, G. Brajušković, V. Rakočević-Stojanović and D. Savić-Pavićević (2018). "Repeat Interruptions Modify Age at Onset in Myotonic Dystrophy Type 1 by Stabilizing DMPK Expansions in Somatic Cells." <u>Frontiers in genetics</u> **9**: 601-601.

Petruska, J., N. Arnheim and M. F. Goodman (1996). "Stability of intrastrand hairpin structures formed by the CAG/CTG class of DNA triplet repeats associated with neurological diseases." <u>Nucleic acids research</u> **24**(11): 1992-1998.

Philips, A. V., L. T. Timchenko and T. A. Cooper (1998). "Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy." <u>Science</u> **280**(5364): 737-741.

Pieretti, M., F. P. Zhang, Y. H. Fu, S. T. Warren, B. A. Oostra, C. T. Caskey and D. L. Nelson (1991). "Absence of expression of the FMR-1 gene in fragile X syndrome." <u>Cell</u> **66**(4): 817-822.

Pinto, R. M., E. Dragileva, A. Kirby, A. Lloret, E. Lopez, J. St Claire, G. B. Panigrahi, C. Hou, K. Holloway, T. Gillis, J. R. Guide, P. E. Cohen, G. M. Li, C. E. Pearson, M. J. Daly and V. C. Wheeler (2013). "Mismatch repair genes Mlh1 and Mlh3 modify CAG instability in Huntington's disease mice: genome-wide and candidate approaches." <u>PLoS Genet</u> **9**(10): e1003930.

Pluciennik, A., R. R. Iyer, M. Napierala, J. E. Larson, M. Filutowicz and R. D. Wells (2002). "Long CTG.CAG repeats from myotonic dystrophy are preferred sites for intermolecular recombination." J Biol Chem **277**(37): 34074-34086.

Polleys, E. J., N. C. M. House and C. H. Freudenreich (2017). "Role of recombination and replication fork restart in repeat instability." <u>DNA Repair</u> **56**: 156-165.

Polling, S. and et al. (2019). "Polyalanine expansions drive a shift into α -helical clusters without amyloid-fibril formation. - PubMed - NCBI." <u>Nat Struct Mol Biol</u>.

Pook, M. A. (2012). "DNA methylation and trinucleotide repeat expansion diseases." <u>DNA</u> <u>Methylation-From Genomics to Technology</u>: 193.

Potaman, V., E. Oussatcheva, Y. L. Lyubchenko, L. Shlyakhtenko, S. Bidichandani, T. Ashizawa and R. Sinden (2004). "Length-dependent structure formation in Friedreich ataxia (GAA) n·(TTC) n repeats at neutral pH." <u>Nucleic acids research</u> **32**(3): 1224-1231.

Potaman, V. N., J. J. Bissler, V. I. Hashem, E. A. Oussatcheva, L. Lu, L. S. Shlyakhtenko, Y. L. Lyubchenko, T. Matsuura, T. Ashizawa, M. Leffak, C. J. Benham and R. R. Sinden (2003). "Unpaired Structures in SCA10 (ATTCT)n·(AGAAT)n Repeats." <u>Journal of Molecular Biology</u> **326**(4): 1095-1111.

Prak, E. T. L. and H. H. J. N. R. G. Kazazian Jr (2000). "Mobile elements and the human genome." <u>Nature Reviews Genetics</u> **1**(2): 134.

Pratte, A., C. Prévost, J. Puymirat and J. Mathieu (2015). "Anticipation in myotonic dystrophy type 1 parents with small CTG expansions." <u>American Journal of Medical Genetics Part A</u> **167**(4): 708-714.

Pretto, D., C. M. Yrigollen, H.-T. Tang, J. Williamson, G. Espinal, C. K. Iwahashi, B. Durbin-Johnson, R. J. Hagerman, P. J. Hagerman and F. J. F. i. g. Tassone (2014). "Clinical and molecular implications of mosaicism in FMR1 full mutations." **5**: 318.

Pretto, D. I., G. Mendoza-Morales, J. Lo, R. Cao, A. Hadd, G. J. Latham, B. Durbin-Johnson, R. Hagerman and F. J. J. o. m. g. Tassone (2014). "CGG allele size somatic mosaicism and methylation in FMR1 premutation alleles." **51**(5): 309-318.

Primerano, B., F. Tassone, R. J. Hagerman, P. Hagerman, F. Amaldi and C. J. R. Bagni (2002). "Reduced FMR1 mRNA translation efficiency in fragile X patients with premutations." **8**(12): 1482-1488.

Pringsheim, T., K. Wiltshire, L. Day, J. Dykeman, T. Steeves and N. Jette (2012). "The incidence and prevalence of Huntington's disease: a systematic review and meta-analysis." <u>Movement Disorders</u> **27**(9): 1083-1091 %@ 0885-3185.

Provenzano, C., M. Cappella, R. Valaperta, R. Cardani, G. Meola, F. Martelli, B. Cardinali and G. Falcone (2017). "CRISPR/Cas9-Mediated Deletion of CTG Expansions Recovers Normal Phenotype in Myogenic Cells Derived from Myotonic Dystrophy 1 Patients." <u>Molecular Therapy</u> - <u>Nucleic Acids</u> **9**: 337-348.

Rajaratnam, A., J. Shergill, M. Salcedo-Arellano, W. Saldarriaga, X. Duan and R. J. F. Hagerman (2017). "Fragile X syndrome and fragile X-associated disorders." **6**.

Ramdzan, Y. M., M. M. Trubetskov, A. R. Ormsby, E. A. Newcombe, X. Sui, M. J. Tobin, M. N. Bongiovanni, S. L. Gras, G. Dewson and J. M. L. Miller (2017). "Huntingtin inclusions trigger cellular quiescence, deactivate apoptosis, and lead to delayed necrosis." <u>Cell reports</u> **19**(5): 919-927 %@ 2211-1247.

Ranum, L., M. Moseley, M. Leppert, G. van den Engh, A. La Spada, M. Koob and J. Day (1999). <u>Massive CTG expansions and deletions may reduce penetrance of spinocerebellar</u>

<u>ataxia type 8</u>. AMERICAN JOURNAL OF HUMAN GENETICS, UNIV CHICAGO PRESS 5720 SOUTH WOODLAWN AVE, CHICAGO, IL 60637-1603 USA.

Rau, F., F. Freyermuth, C. Fugier, J. P. Villemin, M. C. Fischer, B. Jost, D. Dembele, G. Gourdon, A. Nicole, D. Duboc, K. Wahbi, J. W. Day, H. Fujimura, M. P. Takahashi, D. Auboeuf, N. Dreumont, D. Furling and N. Charlet-Berguerand (2011). "Misregulation of miR-1 processing is associated with heart defects in myotonic dystrophy." <u>Nat Struct Mol Biol</u> **18**(7): 840-845.

Ravel-Chapuis, A., G. Bélanger, R. S. Yadava, M. S. Mahadevan, L. DesGroseillers, J. Côté and B. J. Jasmin (2012). "The RNA-binding protein Staufen1 is increased in DM1 skeletal muscle and promotes alternative pre-mRNA splicing." <u>J Cell Biol</u> **196**(6): 699-712.

Reddy, K., M. Tam, R. P. Bowater, M. Barber, M. Tomlinson, K. Nichol Edamura, Y. H. Wang and C. E. Pearson (2011). "Determinants of R-loop formation at convergent bidirectionally transcribed trinucleotide repeats." <u>Nucleic Acids Res</u> **39**(5): 1749-1762.

Reddy, M. S., M. B. Vaze, K. Madhusudan and K. Muniyappa (2000). "Binding of SSB and RecA protein to DNA-containing stem loop structures: SSB ensures the polarity of RecA polymerization on single-stranded DNA." <u>Biochemistry</u> **39**(46): 14250-14262.

Reetz, K., I. Dogan, A. S. Costa, M. Dafotakis, K. Fedosov, P. Giunti, M. H. Parkinson, M. G. Sweeney, C. Mariotti and M. J. T. L. N. Panzeri (2015). "Biological and clinical characteristics of the European Friedreich's Ataxia Consortium for Translational Studies (EFACTS) cohort: a cross-sectional analysis of baseline data." **14**(2): 174-182.

Refsland, E. W. and D. M. Livingston (2005). "Interactions Among DNA Ligase I, the Flap Endonuclease and Proliferating Cell Nuclear Antigen in the Expansion and Contraction of CAG Repeat Tracts in Yeast." <u>Genetics</u> **171**(3): 923.

Reyniers, E., J. J. Martin, P. Cras, E. Van Marck, I. Handig, H. Z. Jorens, B. A. Oostra, R. F. Kooy and P. J. Willems (1999). "Postmortem examination of two fragile X brothers with an FMR1 full mutation." American journal of medical genetics **84**(3): 245-249.

Rhodes, D. and H. J. Lipps (2015). "G-quadruplexes and their regulatory roles in biology." <u>Nucleic Acids Research</u> **43**(18): 8627-8637.

Richard, G.-F. (2015). "Shortening trinucleotide repeats using highly specific endonucleases: a possible approach to gene therapy?" <u>Trends in Genetics</u> **31**(4): 177-186.

Richard, G.-F., B. Dujon and J. Haber (1999). "Double-strand break repair can lead to high frequencies of deletions within short CAG/CTG trinucleotide repeats." <u>Molecular and General Genetics MGG</u> **261**(4-5): 871-882.

Richard, G.-F., A. Kerrest and B. J. M. M. B. R. Dujon (2008). "Comparative genomics and molecular dynamics of DNA repeats in eukaryotes." <u>Microbiol. Mol. Biol. Rev.</u> **72**(4): 686-727.

Richard, G.-F., D. Viterbo, V. Khanna, V. Mosbach, L. Castelain and B. Dujon (2014). "Highly specific contractions of a single CAG/CTG trinucleotide repeat by TALEN in yeast." <u>PLoS One</u> **9**(4): e95611.

Richard, G. F., G. M. Goellner, C. T. McMurray and J. E. Haber (2000). "Recombinationinduced CAG trinucleotide repeat expansions in yeast involve the MRE11–RAD50–XRS2 complex." <u>The EMBO journal</u> **19**(10): 2381-2390. Richard, P. and J. L. J. J. o. m. b. Manley (2017). "R loops and links to human disease." <u>Journal</u> <u>of molecular biology</u> **429**(21): 3168-3180.

Riley, B. E., Y. Xu, H. Y. Zoghbi and H. T. Orr (2004). "The effects of the polyglutamine repeat protein ataxin-1 on the UbL-UBA protein A1Up." Journal of Biological Chemistry **279**(40): 42290-42301 %@ 40021-49258.

Robertson, K. D. (2005). "DNA methylation and human disease." <u>Nat Rev Genet</u> **6**(8): 597-610.

Robinson, D. O., S. R. Hammans, S. P. Read and J. Sillibourne (2005). "Oculopharyngeal muscular dystrophy (OPMD): analysis of the PABPN1 gene expansion sequence in 86 patients reveals 13 different expansion types and further evidence for unequal recombination as the mutational mechanism." <u>Hum Genet</u> **116**(4): 267-271.

Rolfsmeier, M. L., M. J. Dixon, L. Pessoa-Brandão, R. Pelletier, J. J. Miret and R. S. Lahue (2001). "Cis-Elements Governing Trinucleotide Repeat Instability in Saccharomyces cerevisiae." <u>Genetics</u> **157**(4): 1569.

Roy, D., K. Yu and M. R. Lieber (2008). "Mechanism of R-loop formation at immunoglobulin class switch sequences." <u>Molecular and cellular biology</u> **28**(1): 50-60.

Ruzo, A., G. F. Croft, J. J. Metzger, S. Galgoczi, L. J. Gerber, C. Pellegrini, H. Wang, M. Fenner, S. Tse, A. Marks, C. Nchako and A. H. Brivanlou (2018). "Chromosomal instability during neurogenesis in Huntington's disease." <u>Development</u> **145**(2): dev156844.

Sacca, F., G. Puorro, A. Antenora, A. Marsili, A. Denaro, R. Piro, P. Sorrentino, C. Pane, A. Tessa and V. B. J. P. O. Morra (2011). "A combined nucleic acid and protein analysis in Friedreich ataxia: implications for diagnosis, pathogenesis and clinical trial design." **6**(3): e17627.

Sakamoto, N., J. E. Larson, R. R. Iyer, L. Montermini, M. Pandolfo and R. D. Wells (2001). "GGA·TCC-interrupted Triplets in Long GAA·TTC Repeats Inhibit the Formation of Triplex and Sticky DNA Structures, Alleviate Transcription Inhibition, and Reduce Genetic Instabilities." Journal of Biological Chemistry **276**(29): 27178-27187.

Salinas-Rios, V., B. P. Belotserkovskii and P. C. Hanawalt (2011). "DNA slip-outs cause RNA polymerase II arrest in vitro: potential implications for genetic instability." <u>Nucleic acids</u> research **39**(17): 7444-7454.

Samadashwily, G. M., G. Raca and S. M. Mirkin (1997). "Trinucleotide repeats affect DNA replication in vivo." <u>Nature genetics</u> **17**(3): 298.

Sanchez-Contreras, M. and F. Cardozo-Pelaez (2017). "Age-related length variability of polymorphic CAG repeats." <u>DNA repair</u> **49**: 26-32.

Santoro, M., L. Fontana, M. Masciullo, M. L. E. Bianchi, S. Rossi, E. Leoncini, G. Novelli, A. Botta and G. Silvestri (2015). "Expansion size and presence of CCG/CTC/CGG sequence interruptions in the expanded CTG array are independently associated to hypermethylation at the DMPK locus in myotonic dystrophy type 1 (DM1)." <u>Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease</u> **1852**(12): 2645-2652.

Santoro, M., M. Masciullo, R. Pietrobono, G. Conte, A. Modoni, M. L. Bianchi, V. Rizzo, M. G. Pomponi, G. Tasca, G. Neri and G. Silvestri (2013). "Molecular, clinical, and muscle studies in myotonic dystrophy type 1 (DM1) associated with novel variant CCG expansions." <u>J Neurol</u> **260**(5): 1245-1257.

Santoro, M., M. Masciullo, G. Silvestri, G. Novelli and A. Botta (2017). "Myotonic dystrophy type 1: role of CCG, CTC and CGG interruptions within DMPK alleles in the pathogenesis and molecular diagnosis." <u>Clinical Genetics</u> **92**(4): 355-364.

Santoro, M., M. Masciullo, G. Silvestri, G. Novelli and A. Botta (2017). "Myotonic dystrophy type 1: role of CCG, CTC and CGG interruptions within DMPK alleles in the pathogenesis and molecular diagnosis." <u>Clin Genet</u> **92**(4): 355-364.

Santos-Pereira, J. M. and A. J. N. R. G. Aguilera (2015). "R loops: new modulators of genome dynamics and function." <u>Nature Reviews Genetics</u> **16**(10): 583.

Sanz, L. A. and F. Chédin (2019). "High-resolution, strand-specific R-loop mapping via S9.6-based DNA–RNA immunoprecipitation and high-throughput sequencing." <u>Nature Protocols</u> **14**(6): 1734-1755.

Sato, K., I. Yabe, Y. Fukuda, H. Soma, Y. Nakahara, S. Tsuji and H. Sasaki (2010). "Mapping of Autosomal Dominant Cerebellar Ataxia Without the Pathogenic PPP2R2B Mutation to the Locus for Spinocerebellar Ataxia 12Mapping of Cerebellar Ataxia to SCA12 Locus." JAMA Neurology **67**(10): 1257-1262.

Saudou, F., S. Finkbeiner, D. Devys and M. E. Greenberg (1998). "Huntingtin acts in the nucleus to induce apoptosis but death does not correlate with the formation of intranuclear inclusions." <u>Cell</u> **95**(1): 55-66 %@ 0092-8674.

Saugar, I., M. Á. Ortiz-Bazán and J. A. Tercero (2014). "Tolerating DNA damage during eukaryotic chromosome replication." <u>Experimental Cell Research</u> **329**(1): 170-177.

Saveliev, A., C. Everett, T. Sharpe, Z. Webster and R. J. N. Festenstein (2003). "DNA triplet repeats mediate heterochromatin-protein-1-sensitive variegated gene silencing." **422**(6934): 909.

Savkur, R. S., A. V. Philips and T. A. Cooper (2001). "Aberrant regulation of insulin receptor alternative splicing is associated with insulin resistance in myotonic dystrophy." <u>Nature genetics</u> **29**(1): 40.

Savouret, C., E. Brisson, J. Essers, R. Kanaar, A. Pastink, H. te Riele, C. Junien and G. Gourdon (2003). "CTG repeat instability and size variation timing in DNA repair-deficient mice." <u>The EMBO Journal</u> **22**(9): 2264-2273.

Savouret, C., C. Garcia-Cordier, J. Megret, H. te Riele, C. Junien and G. Gourdon (2004). "MSH2-dependent germinal CTG repeat expansions are produced continuously in spermatogonia from DM1 transgenic mice." <u>Molecular and cellular biology</u> **24**(2): 629-637.

Schiavon, G., S. Furlan, O. Marin and S. Salvatori (2002). "Myotonic dystrophy protein kinase of the cardiac muscle: evaluation using an immunochemical approach." <u>Microsc Res Tech</u> **58**(5): 404-411.

Schmidt, M. H. M. and C. E. Pearson (2016). "Disease-associated repeat instability and mismatch repair." <u>DNA Repair (Amst)</u> **38**: 117-126.

Sergeant, N., B. Sablonniere, S. Schraen-Maschke, A. Ghestem, C. A. Maurage, A. Wattez, P. Vermersch and A. Delacourte (2001). "Dysregulation of human brain microtubuleassociated tau mRNA maturation in myotonic dystrophy type 1." <u>Hum Mol Genet</u> **10**(19): 2143-2155. Seznec, H., O. Agbulut, N. Sergeant, C. Savouret, A. Ghestem, N. Tabti, J. C. Willer, L. Ourth, C. Duros, E. Brisson, C. Fouquet, G. Butler-Browne, A. Delacourte, C. Junien and G. Gourdon (2001). "Mice transgenic for the human myotonic dystrophy region with expanded CTG repeats display muscular and brain abnormalities." <u>Hum Mol Genet</u> **10**(23): 2717-2726.

Seznec, H., A. S. Lia-Baldini, C. Duros, C. Fouquet, C. Lacroix, H. Hofmann-Radvanyi, C. Junien and G. Gourdon (2000). "Transgenic mice carrying large human genomic sequences with expanded CTG repeat mimic closely the DM CTG repeat intergenerational and somatic instability." <u>Hum Mol Genet</u> **9**(8): 1185-1194.

Shah, K. A. and S. M. Mirkin (2015). "The hidden side of unstable DNA repeats: Mutagenesis at a distance." DNA repair **32**: 106-112.

Shishkin, A. A., I. Voineagu, R. Matera, N. Cherng, B. T. Chernet, M. M. Krasilnikova, V. Narayanan, K. S. Lobachev and S. M. Mirkin (2009). "Large-scale expansions of Friedreich's ataxia GAA repeats in yeast." <u>Molecular cell</u> **35**(1): 82-92.

Siboni, Ruth B., M. Nakamori, Stacey D. Wagner, Adam J. Struck, Leslie A. Coonrod, Shanee A. Harriott, Daniel M. Cass, Matthew K. Tanner and J. A. Berglund (2015). "Actinomycin D Specifically Reduces Expanded CUG Repeat RNA in Myotonic Dystrophy Models." <u>Cell Reports</u> **13**(11): 2386-2394.

Sicot, G., G. Gourdon and M. Gomes-Pereira (2011). "Myotonic dystrophy, when simple repeats reveal complex pathogenic entities: new findings and future challenges." <u>Human</u> <u>Molecular Genetics</u> **20**(R2): R116-R123.

Sicot, G., L. Servais, D. M. Dinca, A. Leroy, C. Prigogine, F. Medja, S. O. Braz, A. Huguet-Lachon, C. Chhuon, A. Nicole, N. Gueriba, R. Oliveira, B. Dan, D. Furling, M. S. Swanson, I. C. Guerrera, G. Cheron, G. Gourdon and M. Gomes-Pereira (2017). "Downregulation of the Glial GLT1 Glutamate Transporter and Purkinje Cell Dysfunction in a Mouse Model of Myotonic Dystrophy." <u>Cell Rep</u> **19**(13): 2718-2729.

Silveira, I., I. Alonso, L. Guimaraes, P. Mendonca, C. Santos, P. Maciel, J. F. De Matos, M. Costa, C. Barbot and A. J. T. A. J. o. H. G. Tuna (2000). "High germinal instability of the (CTG) n at the SCA8 locus of both expanded and normal alleles." **66**(3): 830-840.

Simone, R., P. Fratta, S. Neidle, G. N. Parkinson and A. M. Isaacs (2015). "G-quadruplexes: emerging roles in neurodegenerative diseases and the non-coding transcriptome." <u>FEBS</u> <u>letters</u> **589**(14): 1653-1668.

Skinner, P. J., B. T. Koshy, C. J. Cummings, I. A. Klement, K. Helin, A. Servadio, H. Y. Zoghbi and H. T. Orr (1997). "Ataxin-1 with an expanded glutamine tract alters nuclear matrix-associated structures." <u>Nature</u> **389**(6654): 971 %@ 1476-4687.

Skourti-Stathaki, K., K. Kamieniarz-Gdula and N. J. Proudfoot (2014). "R-loops induce repressive chromatin marks over mammalian gene terminators." <u>Nature</u> **516**(7531): 436-439.

Skourti-Stathaki, K. and N. J. Proudfoot (2014). "A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression." <u>Genes Dev</u> **28**(13): 1384-1396.

Sobczak, K., T. M. Wheeler, W. Wang and C. A. Thornton (2013). "RNA interference targeting CUG repeats in a mouse model of myotonic dystrophy." <u>Molecular therapy : the journal of the American Society of Gene Therapy</u> **21**(2): 380-387.

Sollier, J. and K. A. J. T. i. c. b. Cimprich (2015). "Breaking bad: R-loops and genome integrity." <u>Trends in cell biology</u> **25**(9): 514-522.

Souidi, A., M. Zmojdzian and K. Jagla (2018). "Dissecting Pathogenetic Mechanisms and Therapeutic Strategies in Drosophila Models of Myotonic Dystrophy Type 1." <u>International journal of molecular sciences</u> **19**(12): 4104.

Spiro, C. and C. T. McMurray (2003). "Nuclease-deficient FEN-1 blocks Rad51/BRCA1mediated repair and causes trinucleotide repeat instability." <u>Molecular and cellular biology</u> **23**(17): 6063-6074.

Spiro, C., R. Pelletier, M. L. Rolfsmeier, M. J. Dixon, R. S. Lahue, G. Gupta, M. S. Park, X. Chen, S. S. Mariappan and C. T. McMurray (1999). "Inhibition of FEN-1 processing by DNA secondary structure at trinucleotide repeats." <u>Molecular cell</u> **4**(6): 1079-1085.

Srivastava, A. K., S. Choudhry, M. S. Gopinath, S. Roy, M. Tripathi, S. K. Brahmachari and S. J. A. o. n. Jain (2001). "Molecular and clinical correlation in five Indian families with spinocerebellar ataxia 12." **50**(6): 796-800.

Steinert, H. (1909). "Myopathologische Beiträge." Journal of Neurology 37(1): 58-104.

Stevanoni, M., E. Palumbo and A. Russo (2016). "The replication of frataxin gene is assured by activation of dormant origins in the presence of a GAA-repeat expansion." <u>PLoS genetics</u> **12**(7): e1006201.

Su, X. A. and C. H. Freudenreich (2017). "Cytosine deamination and base excision repair cause R-loop-induced CAG repeat fragility and instability in Saccharomyces cerevisiae." <u>Proc</u> <u>Natl Acad Sci U S A</u> **114**(40): E8392-e8401.

Subramanian, S., V. M. Madgula, R. George, R. K. Mishra, M. W. Pandit, C. S. Kumar and L. J. B. Singh (2003). "Triplet repeats in human genome: distribution and their association with genes and other genomic regions." **19**(5): 549-552.

Subramanian, S., V. M. Madgula, R. George, R. K. Mishra, M. W. Pandit, C. S. Kumar and L. J. B. Singh (2003). "Triplet repeats in human genome: distribution and their association with genes and other genomic regions." <u>Bioinformatics</u> **19**(5): 549-552.

Subramanian, S., R. K. Mishra and L. J. G. b. Singh (2003). "Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions." <u>Genome biology</u> **4**(2): R13.

Sudmant, P. H., J. O. Kitzman, F. Antonacci, C. Alkan, M. Malig, A. Tsalenko, N. Sampas, L. Bruhn, J. Shendure and E. E. J. S. Eichler (2010). "Diversity of human copy number variation and multicopy genes." <u>Science</u> **330**(6004): 641-646.

Sun, J. H., L. Zhou, D. J. Emerson, S. A. Phyo, K. R. Titus, W. Gong, T. G. Gilgenast, J. A. Beagan, B. L. Davidson and F. Tassone (2018). "Disease-associated short tandem repeats co-localize with chromatin domain boundaries." <u>Cell</u> **175**(1): 224-238. e215.

Sundararajan, R., L. Gellon, R. M. Zunder and C. H. Freudenreich (2010). "Double-strand break repair pathways protect against CAG/CTG repeat expansions, contractions and repeatmediated chromosomal fragility in Saccharomyces cerevisiae." <u>Genetics</u> **184**(1): 65-77.

Sundquist, W. I. and A. Klug (1989). "Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops." <u>Nature</u> **342**(6251): 825-829.

Sutcliffe, J. S., D. L. Nelson, F. Zhang, M. Pieretti, C. T. Caskey, D. Saxe and S. T. Warren (1992). "DNA methylation represses FMR-1 transcription in fragile X syndrome." <u>Hum Mol</u> <u>Genet</u> **1**(6): 397-400.

Sutherland, G. R., E. Baker and R. I. Richards (1998). "Fragile sites still breaking." <u>Trends in</u> <u>Genetics</u> **14**(12): 501-506.

Sweatt, J. D., M. Meaney, E. Nestler and S. Akbarian (2013). "An overview of the molecular basis of epigenetics." <u>Epigenetic Regulation in the Nervous System—Basic Mechanisms and</u> <u>Clinical Impact; Elsevier: San Diego, CA, USA</u>: 3-33.

Szwarocka, S. T., P. Staczek and P. Parniewski (2007). "Chromosomal model for analysis of a long CTG/CAG tract stability in wild-type Escherichia coli and its nucleotide excision repair mutants." <u>Can J Microbiol</u> **53**(7): 860-868.

Tallaksen-Greene, S. J., J. M. Ordway, A. B. Crouse, W. S. Jackson, P. J. Detloff and R. L. Albin (2003). "Hprt (CAG) 146 mice: Age of onset of behavioral abnormalities, time course of neuronal intranuclear inclusion accumulation, neurotransmitter marker alterations, mitochondrial function markers, and susceptibility to 1-methyl-4-phenyl-1, 2, 3, 6-tetrahydropyridine." Journal of Comparative Neurology **465**(2): 205-219 %@ 0021-9967.

Tamanini, F., R. Willemsen, L. van Unen, C. Bontekoe, H. Galjaard, B. A. Oostra and A. T. Hoogeveen (1997). "Differential expression of FMR1, FXR1 and FXR2 proteins in human brain and testis." <u>Hum Mol Genet</u> **6**(8): 1315-1322.

Taneja, K. L., M. McCurrach, M. Schalling, D. Housman and R. H. Singer (1995). "Foci of trinucleotide repeat transcripts in nuclei of myotonic dystrophy cells and tissues." <u>J Cell Biol</u> **128**(6): 995-1002.

Tang, Z. Z., V. Yarotskyy, L. Wei, K. Sobczak, M. Nakamori, K. Eichinger, R. T. Moxley, R. T. Dirksen and C. A. Thornton (2012). "Muscle weakness in myotonic dystrophy associated with misregulated splicing and altered gating of Ca(V)1.1 calcium channel." <u>Hum Mol Genet</u> **21**(6): 1312-1324.

Tassone, F., K. P. Iong, T.-H. Tong, J. Lo, L. W. Gane, E. Berry-Kravis, D. Nguyen, L. Y. Mu, J. Laffin and D. B. J. G. m. Bailey (2013). "FMR1 CGG allele size and prevalence ascertained through newborn screening in the United States." **4**(12): 100.

Tassone, F., C. Iwahashi and P. J. J. R. b. Hagerman (2004). "FMR1 RNA within the intranuclear inclusions of fragile X-associated tremor/ataxia syndrome (FXTAS)." **1**(2): 103-105.

Taylor, A. K., F. Tassone, P. N. Dyer, S. M. Hersch, J. B. Harris, W. T. Greenough and R. J. J. A. j. o. m. g. Hagerman (1999). "Tissue heterogeneity of the FMR1 mutation in a high-functioning male with fragile X syndrome." **84**(3): 233-239.

Telenius, H., H. P. H. Kremer, J. Thellmann, S. E. Andrew, E. Almqvist, M. Anvret, C. Greenberg, J. Greenberg, G. Lucotte, F. Squltierl, A. Starr, Y. P. Goldberg and M. R. Hayden (1993). "Molecular analysis of juvenile Huntington disease: the major influence on (CAG)n repeat length is the sex of the affected parent." <u>Human Molecular Genetics</u> **2**(10): 1535-1540 %@ 0964-6906.

Teng, Y., S. Pramanik, H. Tateishi-Karimata, T. Ohyama and N. Sugimoto (2018). "Drastic stability change of X-X mismatch in d(CXG) trinucleotide repeat disorders under molecular crowding condition." <u>Biochemical and Biophysical Research Communications</u> **496**(2): 601-607.

Theadom, A., M. Rodrigues, R. Roxburgh, S. Balalla, C. Higgins, R. Bhattacharjee, K. Jones, R. Krishnamurthi and V. Feigin (2014). "Prevalence of muscular dystrophies: a systematic literature review." <u>Neuroepidemiology</u> **43**(3-4): 259-268.

Thornton, C. A., E. Wang and E. M. Carrell (2017). "Myotonic dystrophy: approach to therapy." <u>Current Opinion in Genetics & Development</u> **44**: 135-140.

Timchenko, L., N. Timchenko, C. Caskey and R. Roberts (1996). "Novel proteins with binding specificity for DNA CTG repeats and RNA CUG repeats: implications for myotonic dystrophy." <u>Human molecular genetics</u> **5**(1): 115-121.

Timchenko, N. A., Z.-J. Cai, A. L. Welm, S. Reddy, T. Ashizawa and L. T. Timchenko (2001). "RNA CUG repeats sequester CUGBP1 and alter protein levels and activity of CUGBP1." Journal of Biological Chemistry **276**(11): 7820-7826.

Tomé, S., E. Dandelot, C. Dogan, A. Bertrand, D. Geneviève, Y. Péréon, D. M. c. s. group, M. Simon, J.-P. Bonnefont, G. Bassez and G. Gourdon (2018). "Unusual association of a unique CAG interruption in 5' of DM1 CTG repeats with intergenerational contractions and low somatic mosaicism." <u>Human Mutation</u> **39**(7): 970-982.

Tome, S., E. Dandelot, C. Dogan, A. Bertrand, D. Genevieve, Y. Pereon, M. Simon, J. P. Bonnefont, G. Bassez and G. Gourdon (2018). "Unusual association of a unique CAG interruption in 5' of DM1 CTG repeats with intergenerational contractions and low somatic mosaicism." <u>Hum Mutat</u> **39**(7): 970-982.

Tomé, S., I. Holt, W. Edelmann, G. E. Morris, A. Munnich, C. E. Pearson and G. Gourdon (2009). "MSH2 ATPase domain mutation affects CTG*CAG repeat instability in transgenic mice." <u>PLoS genetics</u> **5**(5): e1000482-e1000482.

Tome, S., K. Manley, J. P. Simard, G. W. Clark, M. M. Slean, M. Swami, P. F. Shelbourne, E. R. Tillier, D. G. Monckton, A. Messer and C. E. Pearson (2013). "MSH3 polymorphisms and protein levels affect CAG repeat instability in Huntington's disease mice." <u>PLoS Genet</u> **9**(2): e1003280.

Tomé, S., A. Nicole, M. Gomes-Pereira and G. Gourdon (2014). "Non-radioactive detection of trinucleotide repeat size variability." <u>PLoS currents</u> **6**.

Tomé, S., G. B. Panigrahi, A. López Castel, L. Foiry, D. W. Melton, G. Gourdon and C. E. Pearson (2011). "Maternal germline-specific effect of DNA ligase I on CTG/CAG instability." <u>Human molecular genetics</u> **20**(11): 2131-2143.

Tóth, G., Z. Gáspári and J. J. G. r. Jurka (2000). "Microsatellites in different eukaryotic genomes: survey and analysis." **10**(7): 967-981.

Tremblay, J. P., P. Chapdelaine and J. Rousseau (2019). CRISPR-based methods and products for increasing frataxin levels and uses thereof, Google Patents.

Trochet, D., L. de Pontual, B. Keren, A. Munnich, M. Vekemans, S. Lyonnet and J. Amiel (2007). "Polyalanine expansions might not result from unequal crossing-over." <u>Hum Mutat</u> **28**(10): 1043-1044.

Usdin, K., N. C. House and C. H. Freudenreich (2015). "Repeat instability during DNA repair: Insights from model systems." <u>Critical reviews in biochemistry and molecular biology</u> **50**(2): 142-167.

Usdin, K., N. C. House, C. H. J. C. r. i. b. Freudenreich and m. biology (2015). "Repeat instability during DNA repair: insights from model systems." **50**(2): 142-167.

Utsch, B., K. Becker, D. Brock, M. J. Lentze, F. Bidlingmaier and M. Ludwig (2002). "A novel stable polyalanine [poly(A)] expansion in the HOXA13 gene associated with hand-foot-genital syndrome: proper function of poly(A)-harbouring transcription factors depends on a critical repeat length?" <u>Hum Genet</u> **110**(5): 488-494.

Utsch, B., C. D. McCabe, K. Galbraith, R. Gonzalez, M. Born, J. Dötsch, M. Ludwig, H. Reutter and J. W. Innis (2007). "Molecular characterization of HOXA13 polyalanine expansion proteins in hand–foot–genital syndrome." <u>American Journal of Medical Genetics Part A</u> **143A**(24): 3161-3168.

van Agtmaal, E. L., L. M. André, M. Willemse, S. A. Cumming, I. D. G. van Kessel, W. J. A. A. van den Broek, G. Gourdon, D. Furling, V. Mouly, D. G. Monckton, D. G. Wansink and B. Wieringa (2017). "CRISPR/Cas9-Induced (CTG·CAG)n Repeat Instability in the Myotonic Dystrophy Type 1 Locus: Implications for Therapeutic Genome Editing." <u>Molecular Therapy</u> **25**(1): 24-43.

van den Broek, W. J., M. R. Nelen, G. W. van der Heijden, D. G. Wansink and B. Wieringa (2006). "Fen1 does not control somatic hypermutability of the (CTG) n·(CAG) n repeat in a knock-in mouse model for DM1." <u>FEBS letters</u> **580**(22): 5208-5214.

van den Broek, W. J. A. A., M. R. Nelen, D. G. Wansink, M. M. Coerwinkel, H. te Riele, P. J. T. A. Groenen and B. Wieringa (2002). "Somatic expansion behaviour of the (CTG)n repeat in myotonic dystrophy knock-in mice is differentially affected by Msh3 and Msh6 mismatch–repair proteins." <u>Human Molecular Genetics</u> **11**(2): 191-198.

Van Dijk, J. G., E. A. Van der Velde, R. A. C. Roos and G. W. Bruyn (1986). "Juvenile Huntington disease." <u>Human genetics</u> **73**(3): 235-239 %@ 0340-6717.

Vankan, P. J. J. o. n. (2013). "Prevalence gradients of Friedreich's Ataxia and R1b haplotype in Europe co-localize, suggesting a common Palaeolithic origin in the Franco-Cantabrian ice age refuge." **126**: 11-20.

Verkerk, A. J., M. Pieretti, J. S. Sutcliffe, Y.-H. Fu, D. P. Kuhl, A. Pizzuti, O. Reiner, S. Richards, M. F. Victoria and F. J. C. Zhang (1991). "Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome." <u>Cell</u> **65**(5): 905-914.

Voineagu, I., V. Narayanan, K. S. Lobachev and S. M. Mirkin (2008). "Replication stalling at unstable inverted repeats: Interplay between DNA hairpins and fork stabilizing proteins." <u>Proceedings of the National Academy of Sciences of the United States of America</u> **105**(29): 9936-9941.

Wakimoto, H., C. T. Maguire, M. C. Sherwood, M. M. Vargas, P. S. Sarkar, J. Han, S. Reddy and C. I. Berul (2002). "Characterization of Cardiac Conduction System Abnormalities in Mice with Targeted Disruption of Six5 Gene." <u>Journal of Interventional Cardiac Electrophysiology</u> **7**(2): 127-135.

Wang, E. T., D. Treacy, K. Eichinger, A. Struck, J. Estabrook, H. Olafson, T. T. Wang, K. Bhatt, T. Westbrook and S. Sedehizadeh (2018). "Transcriptome alterations in myotonic dystrophy skeletal muscle and heart." <u>Human molecular genetics</u>.

Wang, G. S., D. L. Kearney, M. De Biasi, G. Taffet and T. A. Cooper (2007). "Elevation of RNAbinding protein CUGBP1 is an early event in an inducible heart-specific mouse model of myotonic dystrophy." <u>J Clin Invest</u> **117**(10): 2802-2811.

Wang, Y., R. M. Pfeiffer, R. Alsaggaf, W. Meeraus, J. C. Gage, L. A. Anderson, R. C. Bremer, N. Nikolenko, H. Lochmuller, M. H. Greene and S. M. Gadalla (2018). "Risk of skin cancer among patients with myotonic dystrophy type 1 based on primary care physician data from the U.K. Clinical Practice Research Datalink." <u>Int J Cancer</u> **142**(6): 1174-1181.

Warby, S. C., A. Montpetit, A. R. Hayden, J. B. Carroll, S. L. Butland, H. Visscher, J. A. Collins, A. Semaka, T. J. Hudson and M. R. Hayden (2009). "CAG expansion in the Huntington disease gene is associated with a specific and targetable predisposing haplogroup." <u>American journal of human genetics</u> **84**(3): 351-366.

Warf, M. B., M. Nakamori, C. M. Matthys, C. A. Thornton and J. A. Berglund (2009). "Pentamidine reverses the splicing defects associated with myotonic dystrophy." <u>Proceedings</u> of the National Academy of Sciences of the United States of America **106**(44): 18551-18556.

Warner, J. P., L. H. Barron, D. Goudie, K. Kelly, D. Dow, D. R. Fitzpatrick and D. J. Brock (1996). "A general method for the detection of large CAG repeat expansions by fluorescent PCR." <u>J Med Genet</u> **33**(12): 1022-1026.

Warren, S. T. (1997). "Polyalanine expansion in synpolydactyly might result from unequal crossing-over of HOXD13." <u>Science</u> **275**(5298): 408-409.

Warren, S. T., Y. Muragaki, S. Mundlos, J. Upton and B. R. Olsen (1997). "Polyalanine Expansion in Synpolydactyly Might Result from Unequal Crossing-Over of HOXD13." <u>Science</u> **275**(5298): 408.

Watase, K., E. J. Weeber, B. Xu, B. Antalffy, L. Yuva-Paylor, K. Hashimoto, M. Kano, R. Atkinson, Y. Sun and D. L. Armstrong (2002). "A long CAG repeat in the mouse Sca1 locus replicates SCA1 features and reveals the impact of protein solubility on selective neurodegeneration." <u>Neuron</u> **34**(6): 905-919 %@ 0896-6273.

Weisman-Shomer, P., Y. Naot and M. Fry (2000). "Tetrahelical forms of the fragile X syndrome expanded sequence d(CGG)(n) are destabilized by two heterogeneous nuclear ribonucleoprotein-related telomeric DNA-binding proteins." J Biol Chem **275**(3): 2231-2238.

Weller, P., A. J. Jeffreys, V. Wilson and A. Blanchetot (1984). "Organization of the human myoglobin gene." <u>The EMBO journal</u> **3**(2): 439-446.

Wheeler, T. M., J. D. Lueck, M. S. Swanson, R. T. Dirksen and C. A. Thornton (2007). "Correction of CIC-1 splicing eliminates chloride channelopathy and myotonia in mouse models of myotonic dystrophy." <u>The Journal of clinical investigation</u> **117**(12): 3952-3957.

Wheeler, V. C., L.-A. Lebel, V. Vrbanac, A. Teed, H. te Riele and M. E. MacDonald (2003). "Mismatch repair gene Msh2 modifies the timing of early disease in HdhQ111 striatum." <u>Human Molecular Genetics</u> **12**(3): 273-281.

Wheeler, V. C., F. Persichetti, S. M. McNeil, J. S. Mysore, S. S. Mysore, M. E. MacDonald, R. H. Myers, J. F. Gusella and N. S. Wexler (2007). "Factors associated with HD CAG repeat instability in Huntington disease." <u>J Med Genet</u> **44**(11): 695-701.

Willis, J. H., G. Isaya, O. Gakh, R. A. Capaldi, M. F. J. M. g. Marusich and metabolism (2008). "Lateral-flow immunoassay for the frataxin protein in Friedreich's ataxia patients and carriers." **94**(4): 491-497.

Woerner, A. C., F. Frottin, D. Hornburg, L. R. Feng, F. Meissner, M. Patra, J. Tatzelt, M. Mann, K. F. Winklhofer and F. U. Hartl (2016). "Cytoplasmic protein aggregates interfere with nucleocytoplasmic transport of protein and RNA." <u>Science</u> **351**(6269): 173-176 %@ 0036-8075.

Wöhrle, D., U. Salat, H. Hameister, W. Vogel and P. Steinbach (2001). "Demethylation, reactivation, and destabilization of human fragile X full-mutation alleles in mouse embryocarcinoma cells." <u>American journal of human genetics</u> **69**(3): 504-515.

Wojtkowiak-Szlachcic, A., K. Taylor, E. Stepniak-Konieczna, L. J. Sznajder, A. Mykowska, J. Sroka, C. A. Thornton and K. Sobczak (2015). "Short antisense-locked nucleic acids (all-LNAs) correct alternative splicing abnormalities in myotonic dystrophy." <u>Nucleic acids research</u> **43**(6): 3318-3331.

Wren, J. D., E. Forgacs, J. W. Fondon III, A. Pertsemlidis, S. Y. Cheng, T. Gallardo, R. Williams, R. V. Shohet, J. D. Minna and H. R. J. T. A. J. o. H. G. Garner (2000). "Repeat polymorphisms within gene regions: phenotypic and evolutionary implications." <u>The American</u> <u>Journal of Human Genetics</u> **67**(2): 345-356.

Wyatt, H. D. M. and S. C. West (2014). "Holliday junction resolvases." <u>Cold Spring Harbor</u> perspectives in biology **6**(9): a023192-a023192.

Xu, M., Y. Lai, J. Torner, Y. Zhang, Z. Zhang and Y. Liu (2014). "Base excision repair of oxidative DNA damage coupled with removal of a CAG repeat hairpin attenuates trinucleotide repeat expansion." <u>Nucleic acids research</u> **42**(6): 3675-3691.

Yadava, R. S., C. D. Frenzel-McCardell, Q. Yu, V. Srinivasan, A. L. Tucker, J. Puymirat, C. A. Thornton, O. W. Prall, R. P. Harvey and M. S. Mahadevan (2008). "RNA toxicity in myotonic muscular dystrophy induces NKX2-5 expression." <u>Nature genetics</u> **40**(1): 61-68.

Yang, Z., R. Lau, J. L. Marcadier, D. Chitayat and C. E. Pearson (2003). "Replication Inhibitors Modulate Instability of an Expanded Trinucleotide Repeat at the Myotonic Dystrophy Type 1 Disease Locus in Human Cells." <u>The American Journal of Human Genetics</u> **73**(5): 1092-1105.

Yanovsky-Dagan, S., M. Avitzour, G. Altarescu, P. Renbaum, T. Eldar-Geva, O. Schonberger, S. Mitrani-Rosenbaum, E. Levy-Lahad, R. Y. Birnbaum, L. Gepstein, S. Epsztejn-Litman and R. Eiges (2015). "Uncovering the Role of Hypermethylation by CTG Expansion in Myotonic Dystrophy Type 1 Using Mutant Human Embryonic Stem Cells." <u>Stem cell reports</u> **5**(2): 221-231.

Yoo, S.-Y., M. E. Pennesi, E. J. Weeber, B. Xu, R. Atkinson, S. Chen, D. L. Armstrong, S. M. Wu, J. D. Sweatt and H. Y. Zoghbi (2003). "SCA7 knockin mice model human SCA7 and reveal gradual accumulation of mutant ataxin-7 in neurons and abnormalities in short-term plasticity." <u>Neuron</u> **37**(3): 383-401 %@ 0896-6273.

Yoon, S.-R., L. Dubeau, M. de Young, N. S. Wexler and N. Arnheim (2003). "Huntington disease expansion mutations in humans can occur before meiosis is completed." <u>Proceedings</u> of the National Academy of Sciences of the United States of America **100**(15): 8834-8838.

Yrigollen, C. M., L. Martorell, B. Durbin-Johnson, M. Naudo, J. Genoves, A. Murgia, R. Polli, L. Zhou, D. Barbouth and A. J. J. o. n. d. Rupchock (2014). "AGG interruptions and maternal age affect FMR1 CGG repeat allele stability during transmission." **6**(1): 24.

Yu, S., M. Pritchard, E. Kremer, M. Lynch, J. Nancarrow, E. Baker, K. Holman, J. Mulley, S. Warren and D. Schlessinger (1991). "Fragile X genotype characterized by an unstable region of DNA." <u>Science</u>: 1179-1181.

Yudkin, D., B. E. Hayward, M. I. Aladjem, D. Kumari and K. Usdin (2014). "Chromosome fragility and the abnormal replication of the FMR1 locus in fragile X syndrome." <u>Human</u> <u>Molecular Genetics</u> **23**(11): 2940-2952.

Yum, K., E. T. Wang and A. Kalsotra (2017). "Myotonic dystrophy: disease repeat range, penetrance, age of onset, and relationship between repeat size and phenotypes." <u>Curr Opin</u> <u>Genet Dev</u> **44**: 30-37.

Zhao, X.-N., D. Kumari, S. Gupta, D. Wu, M. Evanitsky, W. Yang and K. Usdin (2015). "Mutsβ generates both expansions and contractions in a mouse model of the Fragile X-associated disorders." <u>Human molecular genetics</u> **24**(24): 7087-7096.

Zhao, X.-N., R. Lokanga, K. Allette, I. Gazy, D. Wu and K. Usdin (2016). "A MutSβ-Dependent Contribution of MutSα to Repeat Expansions in Fragile X Premutation Mice?" <u>PLoS genetics</u> **12**(7): e1006190-e1006190.

Zhao, X.-N. and K. Usdin (2016). "Ups and Downs: Mechanisms of Repeat Instability in the Fragile X-Related Disorders." <u>Genes</u> **7**(9): 70.

Zhong, N., W. Ju, J. Pietrofesa, D. Wang, C. Dobkin and W. T. Brown (1996). "Fragile X "gray zone" alleles: AGG patterns, expansion risks, and associated haplotypes." <u>Am J Med Genet</u> **64**(2): 261-265.

Zu, T., B. Gibbens, N. S. Doty, M. Gomes-Pereira, A. Huguet, M. D. Stone, J. Margolis, M. Peterson, T. W. Markowski and M. A. J. P. o. t. N. A. o. S. Ingram (2011). "Non-ATG–initiated translation directed by microsatellite expansions." **108**(1): 260-265.



Le tableau ci-après résume l'ensemble des motifs interrompus retrouvés dans les répétitions de triplets CTG chez les patients DM1.

Individus	Age au diagnostic /apparition des symptomes	Myotonie	Cataracte	Autres données diagnostiques	Triplets hérités (n)	Motif d'interruption (5' \rightarrow 3')	Références
Famille A: III-9	62/25	+	+	СМТ	225	(CTG)139(GGC)3G(CCG)20(CCGCTG)14(CTG)35 / (CTG)5(CCGCTG)14(CTG)5	
Famille A: III-16	65/44	+	+	CMT	170		
Famille A: III-17	62/35	+	+	CMT, surdité	179		Braida <i>et al.</i> 2010
Famille A: IV-19	41/20	+	+		213		
Famille A: IV-20	38/24	+	-	CMT,	213	(CTG)n(GGC)3G(CCG)20(CCGCTG)14(CTG)35	
Famille A: IV-21	32/17	-	nd	surdité	220		
Famille A: IV-22	30/24	-	+	-	225		
Famille A-1	Fœtus	na	na	na	230	(CTG)nCTC(CTG)9(CCGCTG)2(CTG)2CCG(CTG)5CCG(C TG)13	
Famille A-4	54/40	+	+	Polineuropathie axonale	600–800	(CTG)nCTC(CTG)9CCG(CTG)5CCG(CTG)5CCG(CTG)13	
Famille A-5	53/42	+		-	450–650	(CTG)nCTC(CTG)9(CCGCTGCTG)5CCG(CTG)10	
Famille A-2	31 /no symptoms	-	-	-	300	(CTG)nCTC(CTG)9(CCGCTG)2(CTG)2CCG(CTG)5CCG(C TG)13	
Famille A3	23 /no symptoms	-	-	-	400–500	(CTG)nCTC(CTG)7CCG(CTG)5CCG(CTG)5CCG(CTG)13	
Famille A6	29/no symptoms	-	-	-	600–750	(CTG)nCTC(CTG)9(CCGCTGCTG)4CCG(CTG)10	
Famille A7	31/no symptoms	-	-	-	270	(CTG)nCTC(CTG)9(CCGCTGCTG)5CCG(CTG)10	Musova <i>et al.</i> 2009
Famille B1	50/40	+	+	-	450	(CTG)n(CCGCTG)33–39(CCG)(CCGCTG)3(CTG)18	
Famille B2	25/pas de symptomes	-	-	-	400	(CTG)n(CCGCTG)35-37(CCG)12CTGCCG(CTG)11	
Famille C	49/23	+		Neuropathie axonale	700	(CTG)n(CCGCTG)2(CCG)8CTG(CCG)6CTG(CCG)6CTGC CGCTG(CCG)2CTG(CCGCTG)3(CCG)2CTG(CCGCTG)3C TG(CCGCTG)4CTG(CCGCTG)4CCG(CCGCTG)3(CTG)3(C CGCTG)2(CTG)10	
Famille D	29/07	+		None	37	(CTG)6(CCGCTG)13(CTG)5	
Famille E		-		Nanisme, skeletal contractures, HCM	43	(CTG)6(CCG CTG)16(CTG)5	

Pt 1	63/47	+	+	Troubles cardiaques, adénome folliculaire	550–700	(CTG)5(CCGCTGCTG)46	
Pt 2	52/47	+	+	Troubles cardiaques, hypothyroïdie	600–830	(CTG)9(CCGCTGCTG)61	
Pt 3	70/30	+	+	Troubles cardiaques et légère neuropathie, hypothyroïdie, diabète	65 (incertain)	(CTG)3(CTG/ <mark>CCG</mark>)5xn	Santoro <i>et al.</i> 2013
Pt 4	32/28	+	-	Troubles cardiaques et neuro- psychologiques, nodules sur la thyroïde	900	(CTG)16 CCG(CTG)4 CCG(CTG)8CCG(CTG)5CCG(CTG)2 CCG(CTG)2 CCG(CTG)8	
Pt 5	78/20	+	+	Troubles cardiaques	970	(CTG)12 CCG(CTG)6 CCG(CTG)5 CCG(CTG)5	
A1	66/58	-	-	nd	1000-1400	(CTG)880-1280(CTG)2CCG(CTG)112CCG(CTG)4	
A2	39/31	+	-		475-640	(CTG)437-602(CTG)14CCG(CTG)18CCG(CTG)4	
A3	Fœtus	na	na		500	(CTG)380(CTG)28CCG(CTG)40CTC(CTG)36CCG(CTG)8C CG(CTG)4	
B1	55/51	+	+		740-930	(CTG)699- 889 CCG (CTG)2(CCG)2(CTG)3(CCGCTG)3(CTG)26	
B2	28/nd	-	-		450-550	(CTG)372- 472(CTG)16CCG(CTG)2(CCGCTG)4CTG(CCGCTG)4CCG (CCGCTG)4CCG(CCGCTG)5(CTG)22	
C1	58/58	+	-		140	(CTG)30(CCG)2(CTG)2CCG(CTG)105	
C2	40/37	+	-		121	(CTG)28(CCG)2(CTG)2CCG(CTG)88	Botta <i>et al.</i> 2017
C3	Fœtus/nd	na	na		113	(CTG)31(CCG)2(CTG)2CCG(CTG)13CCG(CTG)63	
D	38/35	+	-		600-700	(CTG)514-614(CTG)68(CCG)9(CTG)9	
E	56/49	+	-		500-660	(CTG)404-564(CTG)33(CCGCTG)28(CTG)7	
F	70/66	~	-		250	(CTG)208(CTG)5(CCGCTG)16(CTG)5	
G	71/61	+	-		400-580	(CTG)330-510(CTG)8(CCGCTG)17(CTG)2CCG(CTG)25	
н	49/46	+	-		175	(CTG)133(CTG)8CCG(CTG)5(CCG)2CTG(CCG)4(CTG)2(C CG)4CTG(CCG)2(CTG)2CCG(CTG)9	
1	20/15	+	+		260-722	(CTG)188- 650(CTG)CCG(CTG)2CCG(CTG)7CCG(CTG)28CCG(CTG) 7CCG(CTG)22	

DF1-161/39++++Indefange au cardaques520FF </th <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th>								
DF1-2 34/30 + - Hypertrophie du Molet 360 Claig(CCECTG)3(CTG)4(CCECTG)2CTG)2CTGCCC(CTG)17 DF1-3 35/15 + - Hypertrophie du Molet 450 - Hypertrophie du Molet 450 - - Hypertrophie du Molet 360 -		Intolérange au Glucose, Troubles 520 cardiaques 520		+	+	61/39	DF1-1	
DF1-336/15+iNHypertropide under notice450CTG)n(CG)3e(CTG)nCCG(CTG)7CCG(CTG)12Preservice of a particeDF2-145/40+-Toubles cardiaque320CTG)n(CG)3e(CTG)6CCG)3(CTG)6CCG(CTG)7CCG(CTG)8cCG)Preservice of a particeDF2-214/12+-Toubles cardiaque240CTG)n(CG)3(CTG)6(CCG)3(CTG)6CCG)3(CTG)6CCG)(CTG)8cCG)Preservice of a particeDF3-150/45++Hypertryoride, troubles cardiaque300CTG)nCCG(CCGCTG)4CTG)CGCGTG)4CTG(CCGCTG)4CTGPreservice of a particeDF4-1nd-Asymptomatique800CTG)nCCG(CCGCTG)4CTG)CGCGTG)4CTG(CCGCTG)4CTGPreservice of a particeDF5-1nd-Asymptomatique800CTG)nCCG(CGCGTG)4CTG)53.67(TG)53.67Preservice of a particeDMGV1425/naAsymptomatique327CTG)802-030(CCGCTGCTG)10.14(CTG)15.23Preservice of a particeDMGV1522/naAsymptomatique327CTG)31(CA0)(CTG)161.53Preservice of a particeDMGV1522/naAsymptomatique327CTG)31(CA0)(CTG)161.53Preservice of a particeDMGV1425/naPreservice of a particePreservice of a partice-A1A1<	TG)17	(CTG)n(CCGCTG)3(CTG)4(CCGCTG)2CTGCCG(CTG)17	350	Hypertrophie du mollet	34/30 + -		DF1-2	
DF2-145/40+Troubles cardiaques320(CTG)n(CCG)3(CTG)nCCG(CTG)CCG(CTG)12Persouic et al.DF2-214/12+-Troubles cardiaques200(CTG)n(CCG)3(CTG)6(CG)3(CTG)7CCG(CTG)8CCG)122017 et 2018DF3-150/45++Hyperhyroide, Hyperparathyroide300(CTG)n(CCG)3(CTG)6(CG)3(CTG)7CCG(CG)3(CTG)2CCGTG)8CCG)Persouic et al.DF4-139/99++Hyperhyroide, Hyperparathyroide300(CTG)nCCG(CCG)4(CTG)6(CG)3(CTG)2DPersouic et al.DF5-228/22++Toubles cardiaques2500(CTG)10C-CTG)25Persouic et al.DF5-228/22++Toubles cardiaques2500(CTG)10C-CTG)25Persouic et al.DMGV 1425/naAsymptomatique381(CTG)180-240(CCG)CTG)3-67Persouic et al.DMGV1522/naAsymptomatique327(CTG)260-320(CCG)CTG)10-14(CTG)15-23Persouic et al.DMGV16222/naAsymptomatique327(CTG)31(CAG)(CTG)10-14(CTG)15-23Persouic et al.A1Asymptomatique125(CTG)31(CAG)(CTG)10-14(CTG)15-23Persouic et al.A1Asymptomatique125(CTG)31(CAG)(CTG)10-14Persouic et al.A2Asymptomatique125(CTG)30(CAG)(CTG)10-14Persouic et al.A4.10A4.30 <td></td> <td></td> <td>450</td> <td>Hypertrophie du mollet</td> <td>-</td> <td>+</td> <td>35/15</td> <td>DF1-3</td>			450	Hypertrophie du mollet	-	+	35/15	DF1-3
$\begin{array}{ c c c c c c c c c c c c c c c c c c c$			320	Troubles cardiaques	+	+	45/40	DF2-1
DF3-1 50/45 + + Hypethyroidie, troubles cardiaques 240 (CTG)n(CCG)3(CTG)6(CCG)3(CTG)7CCG(CTG)8CCG(CTG)8CCG(CTG) Network DF4-1 39/59 + + Hypethyroidie, troubles cardiaques 300 (CTG)nCCG(CCGCTG)4CTG(CCGCTG)4CTG(CCGCTG)4CTG(CCGCTG)4CTG(CCGCTG)4CTG) DF5-1 nd Image: troubles cardiaques 80 Image: troubles cardiaques 250 (CTG)nCCCG1626-00000000000000000000000000000000000	Pesovic 2017 et		200	Troubles cardiaques	-	+	14/12	DF2-2
DF4-139/59++Hyperhyroidie, Hyperparathyroidie300(CTG)nCCG(CGCTG)4CTG(CCGCTG)4CTG(CCGCTG)4CTG(CCGCTG)4CTG)CCGCTG)4 (CTG)5(CCG)4(CTG)6(CCG)3(CTG)20)DF5-1nd1Asymptomatique80DF5-226/22++Toubles cardiaques250(CTG)nCC(CTG)26UU-Asymptomatique80DMGV 1425/na-Assymptomatique381(CTG)180-240(CCGCTG)53-67(CTG)53-67DMGV1522/naAssymptomatique327(CTG)280-320(CCGCTGCTG)10-14(CTG)15-232019DMGV18234/46++Symptomatique327(CTG)31(CAG)(CTG)138A1Assymptomatique294(CTG)31(CAG)(CTG)118A2FoetusA1-Assymptomatique125(CTG)31(CAG)(CTG)188nanaA3Pas d'anticipation125(CTG)31(CAG)(CTG)108naA4.2-130(CTG)29(CAG)(CTG)106-naA4.3Foetusnana365(CTG)11(CCG)1(CTG)2(CCG)1(CTG)4CG)1(CTG)4CG)1(CTG)4GB3.1Foetusnana300(CTG)11(CCG)1(CTG)2(CCG)1(CTG)4CG)1(CTG)4CG)1(CTG)200B3.3Foetusnana300(CTG)11(CCG)1(CTG)2(CCG)1(CTG)4CG)1(CTG)205B3.3Foetusnana200(CTG)11(CCG)1(CTG)2(CG(CT	(CTG)n(CCG)3(CTG)6(CCG)3(CTG)7CCG(CTG)8CCG(CT G)8	240	Hypothyroïdie, troubles cardiaques	50/45 + +		DF3-1	
DF5-1ndImageAsymptomatique80DF5-226/22++Troubles cardiaques250(CTG)nCTC(CTG)26DMGV 1425/na-Assymptomatique381(CTG)180-240(CCGCTG)53-67(CTG)53-67Demming et al.DMGV1522/naAssymptomatique327(CTG)260-320(CCGCTGC1G)10-14(CTG)15-23Demming et al.DMGV18234/d6++Symptomatique327(CTG)260-320(CCGCTGC1G)10-14(CTG)15-23Demming et al.DMGV18234/d6++Symptomatique327(CTG)31(CAG)(CTG)138Demming et al.A1Symptomatique327(CTG)31(CAG)(CTG)138Demming et al.A2A4.1Asymptomatique-150(CTG)31(CAG)(CTG)18Net Since Sinc	SCTG)2(+ + Hyperthyroïdie, Hyperparathyroïdie 300 (CTG)nCCG(CCGCTG)4CTG(CCGCTG)4CTG(CCGCTG)2(CTG)5(CCG)4(CTG)6(CCG)3(CTG)20		+	39/59	DF4-1		
DF5-2 26/22 + + Troubles cardiaques 250 (CTG)nCTC(CTG)26 DMGV 14 25/na - Assymptomatique 381 (CTG)180-240(CCGCTG)53-67(CTG)53-67 Cumming et al. 2019 DMGV15 22/na - Assymptomatique 327 (CTG)260-320(CCGCTGCTG)10-14(CTG)15-23 Cumming et al. 2019 DMGV182 34/46 + + Symptomatique 327 (CTG)300.CCG(CTG)161-14(CTG)15-23 Cumming et al. 2019 MGV182 34/46 + + Symptomatique 327 (CTG)300.CCG(CTG)161-14(CTG)15-23 Cumming et al. 2019 MA 34/46 + + Symptomatique 327 (CTG)300.CCG(CTG)161-14 2019 A1 - Symptomatique 170 (CTG)31(CAG)(CTG)118 7 7 A2 - - - Pas d'anticipation 1140 (CTG)29(CAG)(CTG)108 7 7 A4.1 - - na na 125 (CTG)30(CAG)(CTG)100 7 7 A4.2 -			80	Asymptomatique	nd		DF5-1	
DMGV 1425/na-Assymptomatique381(CTG)180-240(CCGCTG)53-67(CG)53-67 (CTG)260-320(CCGCTGCTG)10-14(CTG)15-23 2019 $Umming et al.$ 2019DMGV1522/na-Assymptomatique327(CTG)260-320(CCGCTGCTG)10-14(CTG)15-23 (CTG)200-300(CCG)(TG)11-59 $Umming et al.$ 2019DMGV18234/46++Symptomes légers294(CTG)200-300(CCG)(TG)10-14(CTG)15-23 (CTG)200-300(CCG)(TG)11-59 $Umming et al.$ 2019MathematiqueA1A1A1A1A1A1A1A1A1A1A1A1A1A1A1A1A1A1A3A1A1A1A1A1A1A1A2A1A1A1A3A1A3A1A3A1A3A1A1A1A1A1A1 <t< td=""><td></td><td>(CTG)nCTC(CTG)26</td><td>250</td><td>Troubles cardiaques</td><td>+</td><td>+</td><td>26/22</td><td>DF5-2</td></t<>		(CTG)nCTC(CTG)26	250	Troubles cardiaques	+	+	26/22	DF5-2
DMGV 14 25/na - Assymptomatique 381 (CTG)180-240(CCGCTG)53-67(TG)53-67 $umming et al.$ DMGV15 22/na - Assymptomatique 327 (CTG)260-320(CCGCTGCTG)10-14(CTG)15-23 $umming et al.$ DMGV182 34/46 + + Symptomatique 224 (CTG)20-300(CCG)(TG)41-59 $umming et al.$ DMGV182 34/46 + + Symptomatigues 294 (CTG)20-300(CCG)(TG)41-59 $umming et al.$ Mathematical - - - - - $umming et al.$ A1 - - - - - - $umming et al.$ A2 - <								
$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$		(CTG)180-240(CCGCTG)53-67(CTG)53-67	381	Assymptomatique			25/na	DMGV 14
DMGV182 34/46 + Symptômes légers 294 (CTG)200-300(CCG)(CTG)41-59 DMGV182 A1	Cummin 201	(CTG)260-320(CCGCTGCTG)10-14(CTG)15-23	327	Assymptomatique	22/na 34/46 + +		22/na	DMGV15
A1 Image: Norm of the second sec	201	(CTG)200-300(CCG)(CTG)41-59	294	Symptômes légers			34/46	DMGV182
A1(CTG)31(CAG)(CTG)138A2150(CTG)31(CAG)(CTG)18A3A3Approximation150(CTG)29(CAG)(CTG)108A4.1Approximation125(CTG)29(CAG)(CTG)108A4.2130(CTG)29(CAG)(CTG)100(CTG)29(CAG)(CTG)40A4.3FoetusnanaB1Main MainMain Main125(CTG)30(CAG)(CTG)4(CCG)1(CTG)4(CCG)1(CTG)45B2Main MainMain Main130(CTG)11(CCG)1(CTG)2(CCG)1(CTG)4(CCG)1(CTG)45B3.1FoetusnanaB3.2FoetusnanaB3.3FoetusnanaB3.3FoetusnanaMain MainMain250(CTG)11(CCG)1(CTG)2(CCG)1(CTG)4(CCG)1(CTG)250B3.3FoetusnanaB3.4FoetusnanaB3.5FoetusnanaB3.4FoetusnanaB3.4FoetusnaB3.5FoetusnaB3.6CTG)11(CCG)1(CTG)2(CCG)1(CTG)4(CCG)1(CTG)250B3.7FoetusnaB3.7FoetusnaB3.7FoetusnaB3.7FoetusnaB3.7FoetusnaB3.7FoetusnaB3.7FoetusnaB3.7FoetusnaB3.7FoetusnaB3.7FoetusnaB3.7FoetusnaB3.7FoetusnaB3.7Foetus <t< td=""><td></td><td></td><td></td><td></td><td><u>.</u></td><td></td><td><u> </u></td><td></td></t<>					<u>.</u>		<u> </u>	
A2150(CTG)31(CAG)(CTG)118A3AAsymptotic provided to the p		(CTG)31 <mark>(CAG)</mark> (CTG)138	170		_			A1
$ \begin{array}{ c c c c c } \hline A3 & \\ \hline A4.1 & \\ \hline A4.2 & \\ \hline A4.2 & \\ \hline A4.2 & \\ \hline A4.2 & \\ \hline A4.3 & \hline Fectus & na & na & \\ \hline & & & & \\ \hline A4.3 & \hline Fectus & na & na & \\ \hline & & & &$		(CTG)31 <mark>(CAG)(</mark> CTG)118	150	Pas d'anticipation	_			A2
$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$		(CTG)29 <mark>(CAG)</mark> (CTG)108	140		Assymptomatiques			A3
A4.2Image: Constant of the symbol of the symbo		(CTG)29 <mark>(CAG)</mark> (CTG)95	125					A4.1
A4.3Fœtusnana125(CTG)30(CAG)(CTG)94Tomé et al. 2018B1 $$		(CTG)29 <mark>(CAG)</mark> (CTG)100	130					A4.2
B1 365 (CTG)11(CCG)1(CTG)2(CCG)1(CTG)4(CCG)1(CTG)345 B2 310 (CTG)11(CCG)1(CTG)2(CCG)1(CTG)4(CCG)1(CTG)290 B3.1 Fœtus na na B3.2 Fœtus na na B3.3 Fœtus na 10 CTG)11(CCG)1(CTG)2(CCG)1(CTG)4(CCG)1(CTG)280 235 (CTG)11(CCG)1(CTG)2(CCG)1(CTG)4(CCG)1(CTG)215	Tomé et a	(CTG)30 <mark>(CAG)</mark> (CTG)94	125		na	Fœtus na na		A4.3
B2 310 (CTG)11(CCG)1(CTG)2(CCG)1(CTG)4(CCG)1(CTG)290 B3.1 Fœtus na na B3.2 Fœtus na na B3.3 Fœtus na na B3.3 Fœtus na na	3)345	(CTG)11(CCG)1(CTG)2(CCG)1(CTG)4(CCG)1(CTG)345	365					B1
B3.1 Fœtus na na Pas d'anticipation 300 (CTG)11(CCG)1(CTG)2(CCG)1(CTG)4(CCG)1(CTG)280 B3.2 Fœtus na na 235 (CTG)11(CCG)1(CTG)2(CCG)1(CTG)4(CCG)1(CTG)215 B3.3 Fœtus na na 250 (CTG)11(CCG)1(CTG)2(CCG)1(CTG)4(CCG)1(CTG)230	3)290	(CTG)11(CCG)1(CTG)2(CCG)1(CTG)4(CCG)1(CTG)290	310	Pas d'anticipation	FœtusnanaFœtusnanaFœtusnana		B2	
B3.2 Fœtus na na B3.3 Fœtus na na B3.3 Fœtus na na	3)280	(CTG)11(CCG)1(CTG)2(CCG)1(CTG)4(CCG)1(CTG)280	300				B3.1	
B3.3 Fœtus na na 250 (CTG)1(CCG)1(CTG)2(CCG)1(CTG)4(CCG)1(CTG)230	3)215	(CTG)11(CCG)1(CTG)2(CCG)1(CTG)4(CCG)1(CTG)215	235				B3.2	
	3)230	(CTG)11(CCG)1(CTG)2(CCG)1(CTG)4(CCG)1(CTG)230	250				Fœtus	B3.3



La revue ci-après apporte un résumé les différents modèles murins de la maladie de Huntington ayant permis l'analyse des triplets répétés de CAG/CTG en 2017. J'ai co-écrit cette revue avec Stéphanie Tomé.

Genetic Modifiers of CAG.CTG Repeat Instability in

Huntington's Disease Mouse Models

Elodie Dandelot and Stéphanie Tomé

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/66438

Abstract

Huntington's disease (HD) is a dominantly inherited neurodegenerative disorder whose characterstics were first described by George Huntington in 1872. Several decades later, in 1993, the mutation behind this disease was found to be an unstable expanded CAG repeat within exon 1 of the HTT gene localized on the short arm of chromosome 4. The majority of HD patients carry more than 40 CAG repeats, which become unstable and usually increase in size in successive generations and in tissues. In order to dissect the molecular mechanisms underlying CAG repeat instability, several HD mouse models have been created in the 1990s. Significant data have revealed that the absence of proteins from the mismatch repair (MMR) or the base and nucleotide excision repair decreased the pathogenic expansion-biased somatic mosaicism and/or intergenerational expansions. Some polymorphic variants of MMR genes have also been associated with reduced somatic expansions. Since expansionbiased somatic mosaicism likely contributes to disease manifestations, these results suggest that genetic modifiers of instability may also affect disease severity. In this chapter, we provide an overview of the data recently published about DNA instability; the roles of genetic modifiers of trinucleotide repeat dynamics in mouse models; and the possible therapeutic interventions.

Keywords: Huntington disease, DNA instability, mouse models, genetic modifiers, MMR

1. Introduction

Expansions of repetitive DNA sequences, including trinucleotide repeats, are associated with a large number of neurological and neuromuscular disorders, such as fragile X syndrome,



© 2017 The Author(s). Licensee InTech. This chapter is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 2 Huntington's Disease - Molecular Pathogenesis and Current Models

myotonic dystrophy type 1 and Huntington's disease (HD) [1, 2]. In the healthy population, the triplet repeat tract size varies between 5 and 30 repeats and is stable. In HD patients, the pathogenic allele contains more than 40 repeats and becomes highly unstable and usually increases in size in successive generations (intergenerational instability) and in somatic tissues (somatic instability). Longer expanded alleles are associated with more severe forms of disease and result in a decreasing age of onset from one generation to the next [1, 3, 4]. Among trinucleotide repeat disorders, HD disease is the fourth reported.

1.1. Clinical picture of HD

Huntington's disease is an autosomal dominant neurodegenerative disorder with a worldwide incidence varying from 0.1 to 10 per 100,000 people depending on the country. The estimation of prevalence varies according to haplogroups studied: it is estimated from 2 to 7 per 100,000 in the Caucasians and only 0.1–1 per 100,000 in Asians and Africans [5, 6]. Adult-onset Huntington disease is the most common form of HD and usually presents in early middle life. HD symptoms include uncontrolled movements such as chorea, progressive cognitive impairment and neuropsychiatric manifestations. The rare early-onset form of the disease also called juvenile form presents more severe symptoms with rigidity and motor dysfunctions [7]. HD symptoms and severity vary greatly among family patients and between juvenile and adult onset forms. Currently, no treatment is suitable to stop or reverse any form of HD.

1.2. Genetic of HD

HD is caused by an unstable expanded CAG repeat within exon 1 of the *huntingtin* (*HTT*) gene also called *HD* or *IT15* that localizes on the short arm of chromosome four, 4p16.3 [8]. The normal *HTT* gene contains from 5 to 35 stable CAG repeats, while the majority of HD patients have expanded repeats of above 40 CAG units that are fully penetrant. In rare cases, HD symptoms are associated with small CAG repeats from 36 to 39 CAG, which show low penetrance [9, 10]. Abnormal CAG repeat tracts become unstable in the germline, with a striking tendency toward expansions. Because longer alleles are associated with more severe form of HD, expansion-biased intergenerational instability results in a decreasing age of onset from one generation to the next, a phenomenon known as anticipation. Typically, 40–50 CAG repeats correlate with later-onset of HD, whereas a mutation greater than 50 CAG repeats results in a juvenile form. Two large analyses in HD patients (360 and 440 individuals, respectively) have reported a high negative correlation between the disease age of onset and the inherited CAG repeat length [11, 12]. Intergenerational instability biased toward expansions provides the molecular basis for clinical anticipation observed in HD (**Figure 1**).

1.2.1. Intergenerational instability

The frequencies of expanded, unchanged and contracted alleles have been investigated by directly comparing the length of the repeat tract in each parent with that is observed in their progeny to estimate the degree of intergenerational instability in each set of HD cohort. Small normal alleles with CAG repeat size ranging from 10 to 28 CAG are genetically stable with germline mutation rates <1% per generation [13]. However, the mutation frequency rises

Genetic Modifiers of CAG.CTG Repeat Instability in Huntington's Disease Mouse Models 3 http://dx.doi.org/10.5772/66438



Figure 1. CAG repeat dynamics in HD: features and implications of intergenerational and somatic instabilities.

dramatically with the increasing size of the allele. Indeed, a CAG repeat size change on expanded allele in the range of 36–49 repeats occurs in >70% of transmissions from affected parents to HD children. A similar rate of expansion was found between multiethnic cohorts [13–19]. In the two largest cohorts (>250 parent-offspring pairs), the frequency of expansions was estimated to be 52.1% in a multiethnic HD population and 67.3% in the Dutch cohort, whereas only 18.1% and 25.2% contractions were observed, respectively [13, 18]. For individuals carrying more than 49 CAG repeats, the mutation rates go up to >95% per generation [14, 20]. In all cases, the frequency of expansions always exceeds the frequency of contractions in HD populations. The instability of the CAG repeat between generations depends on the sex of the transmitting parent and the length of the repeat itself. Studies of the two cohorts of HD individuals with the mean size of \sim 43 CAG repeats have shown that 61–68% of paternal transmissions resulted in expansions, whereas the majority (>60%) of maternal transmissions resulted in contractions or CAG stabilization [13, 18]. The largest expansions, associated with the juvenile form of HD, are almost observed in male transmissions and are influenced by the CAG repeat length of the transmitting parent [13]. The largest HD cohort study (337 transmissions) has shown that the age of the transmitting parents and the sex of offspring do not affect the intergenerational instability, suggesting that the gender of affected parents is the Genetic Modifiers of CAG.CTG Repeat Instability in Huntington's Disease Mouse Models 3 http://dx.doi.org/10.5772/66438



Figure 1. CAG repeat dynamics in HD: features and implications of intergenerational and somatic instabilities.

dramatically with the increasing size of the allele. Indeed, a CAG repeat size change on expanded allele in the range of 36-49 repeats occurs in >70% of transmissions from affected parents to HD children. A similar rate of expansion was found between multiethnic cohorts [13–19]. In the two largest cohorts (>250 parent-offspring pairs), the frequency of expansions was estimated to be 52.1% in a multiethnic HD population and 67.3% in the Dutch cohort, whereas only 18.1% and 25.2% contractions were observed, respectively [13, 18]. For individuals carrying more than 49 CAG repeats, the mutation rates go up to >95% per generation [14, 20]. In all cases, the frequency of expansions always exceeds the frequency of contractions in HD populations. The instability of the CAG repeat between generations depends on the sex of the transmitting parent and the length of the repeat itself. Studies of the two cohorts of HD individuals with the mean size of \sim 43 CAG repeats have shown that 61–68% of paternal transmissions resulted in expansions, whereas the majority (>60%) of maternal transmissions resulted in contractions or CAG stabilization [13, 18]. The largest expansions, associated with the juvenile form of HD, are almost observed in male transmissions and are influenced by the CAG repeat length of the transmitting parent [13]. The largest HD cohort study (337 transmissions) has shown that the age of the transmitting parents and the sex of offspring do not affect the intergenerational instability, suggesting that the gender of affected parents is the

4 Huntington's Disease - Molecular Pathogenesis and Current Models

major modifier of intergenerational instability [18]. Repeat size variability has been investigated in spermatogonia, postmitotic spermatid and matura spermatozoa collected by laser capture microdissection of testis from two HD patients in order to determine the timing of repeat instability. Interestingly, CAG repeat expansions were already present before the end of the first meiotic division and the frequency continues to increase in postmeiotic cell population suggesting that the primary source of instability occurs in spermatogonia [21].

1.2.2. Somatic instability

Several studies have reported that the expanded CAG repeat allele is also unstable in somatic tissues and increases in length over time [22–25]. Somatic CAG repeat size variation was analyzed by bulk PCR in each tissue whereas the degree of somatic mosaicism was quantified by a more sensitive PCR-based approach called Small-Pool PCR [26]. This method allows to accurately assess the variation of CAG repeat length of each HD expanded allele in tissues, using successive DNA dilutions in order to amplify few template molecules per reaction (Figure 2A). The dynamics of somatic CAG repeat instability varies between and within tissues with the highest instability observed in the striatum and cortex, two tissues that show the most pronounced neuropathological abnormalities [22, 23, 25]. In a large Venezuelan HD cohort, a positive correlation was reported between the size of progenitor alleles (inherited alleles) and the expansion-biased somatic mosaicism in buccal cells from individuals at the same age. This observation suggests that the size of the inherited CAG repeat is an important modulator of somatic instability [27]. Furthermore, it has been reported that CAG repeat expansion length in the cortex is associated with an earlier age of disease onset suggesting that somatic instability is a significant predictor of the age of onset [28]. Interestingly, somatic instability was not observed in two fetuses at 12–13 weeks suggesting that the somatic expansion event occurs later in the stages of fetal development or from birth throughout the patient's life [29].

Together, these data have clearly demonstrated the contribution of the sex of the transmitting parent and the inherited length of the CAG repeat in the dynamics of intergenerational and somatic instability in HD patients. Moreover, both germline and somatic mosaicism level seems to be linked to the disease onset and to the progression of HD symptoms. Thus, aiming at decreasing the size of expanded alleles or the level of somatic mosaicism would be an attractive therapeutic strategy. In the majority of analyses, the degree of expansion length variability between tissues and individuals cannot be explained only by the age, sex of the transmitting parent and the progenitor allele size, therefore implying that genetic factors might influence either germline or somatic instability. In 2012, the study of a large Portuguese HD cohort has reported some HD families with extreme repeat length changes from parents to offspring suggesting the existence of modifiers that may be heritable [19]. Hence, the understanding of CAG repeat instability is crucial to improve the therapeutic possibilities. Analyses of genetic modifiers of instability and dissection of mechanisms involved in this process are compromised by the limited accessibility of human samples and clinical information. Then, knockout, transgenic and knock-in HD mouse models have been generated to dissect the molecular mechanisms of instability and the pathogenesis of HD disease [30, 31].

Genetic Modifiers of CAG.CTG Repeat Instability in Huntington's Disease Mouse Models 5 http://dx.doi.org/10.5772/66438



Figure 2. Methods to analyze CAG repeat length in germline and somatic tissues.

2. Mouse models of CAG repeat instability

The dynamics of expanded CAG repeat has already been analyzed in different simple organism strains such as bacteria and yeast by inserting a plasmid with a pathogenic CAG repeat. Analyses in *E. coli* and *S. cerevisiae* have provided valuable insight into factors affecting the CAG repeat instability. However, these organisms displayed a CAG repeat instability biased toward contractions in clear contrast to HD patients. Furthermore, both these organisms differ from mammals by cellular processes such as DNA repair and replication pathways. Therefore, mouse models have been generated to identify genetic modifiers of instability and to specify the mechanisms by which they act in HD. These mouse models including two HD transgenic mice with short gene fragment or BAC (R6 and BACHD), eight knock-in (the HdhQ20, HdhQ50, HdhQ92, HdhQ111, Hdh4/80, or Hdh6/72 lines, HdhQ150 and HdhQ80), have been created to analyze the dynamic of CAG repeat instability in germline and somatic tissues by different methods [24, 32–36] (Table 1). The first method determines CAG repeat size by using unlabeled primers flanking CAG repeat. PCR and SP-PCR products can be resolved on agarose gel with internal size standards and detected with radioactive probes [37]. The second method measures the length of CAG tracts by using primers flanking CAG repeat expansions, labeled with the 5-carboxyfluoroscein fluorochrome. PCR products are electrophoresed/separated in an automated sequencer together with internal size standards. In this case, the sizing of the PCR fragment is determined using GeneMapper software that represents the PCR fragments by peaks with single repeat unit resolution (Figure 2).

6 Huntington's Disease - Molecular Pathogenesis and Current Models

	Mouse models	Genetic background	Transgene	CAG repeat length	Mutation rate	Intergenerational instability (CAG length variation)	Somatic instability (partial list	References
Transgenic mice	BACHD	FVB	Human <i>HTT</i> locus	97	None	None	None	[34]
	R6 (Excluded R6/T)	CBA/C57BL/6	Human <i>HTT</i> exon 1	>110	65–84%		Striatum> kidney> cerebellum	[32, 45, 46, 51, 52]
Knock-in mice	Hdh ^{Q80}	C57BL6/J		80	~20%	↑ expansions (male transmissions)		[35]
	Hdh4/ ^{Q80}	129Svter/ C57BL6		80	~20%	↓ contractions (female transmissions)	Striatum> cerebellum> liver	[25, 36] >
	Hdh6/ ^{Q72}			72	~20%			
	Hdh ^{Q50}			48	4%	Low	None	[54]
	Hdh ^{Q20}		CAG repeat locus	18	None	None		[24]
	Hdh ^{Q92}	129SvEv/CD1		90	49%	↑ expansions (male transmissions)	Striatum> kidney>	[24, 56, 57]
	Hdh ^{Q111}			109	73%	↓ contractions (female transmissions)	cerebellum	[24, 56, 57, 58]
	Hdh ^{Q150}	C57BL6/129Ola	a	150	16%	ND	Striatum> olfactory bulb> cerebellum	[25, 33, 38, 39]

Table 1. HD mouse models of CAG repeat instability.

BACHD mouse model was established by the introduction of a full-length human htt locus containing exon 1 with 97 mixed CAA-CAG repeats in the FVB background. These mice do not exhibit any repeat instability or contraction in germline and in brain tissues at 12 month of age [34]. The stability of CAG triplet repeat results from the CAA interruption within the CAG repeat tract, which probably modifies the DNA structure and then the repeat dynamics [1]. Compared to BACHD mice, HdhQ150 knock-in mice were generated by replacement of the murine short CAG repeat in exon 1 with a 150 CAG repeat expansion in a mixed C57BL/ 6/129Ola genetic background [33]. HdhQ150 animals reproduce somatic mosaicism in different brain regions, most particularly in the striatum like HD patients [25, 38]. HdhQ150 mice displayed some HD symptoms that seem more severe in homozygous mice [33, 39, 40]. Recently, heterozygous hdhQ250 mice have been generated from hdhQ150 by selective breeding and shown more severe neurological symptoms than heterozygous hdhQ150 mice [41]. Hdh4/Q80 and Hdh6/Q72 mice have also been obtained by replacement of short CAG repeat with 72 or 80 CAG repeat expansions in htt murine gene context. Both these lines have shown intergenerational instability biased toward expansions in paternal transmissions and contractions in maternal transmissions like in HD patients. However, the mutation frequency is only 20% across generations compared to 70% in HD individuals [36]. Hdh4/Q80 and Hdh6/

Genetic Modifiers of CAG.CTG Repeat Instability in Huntington's Disease Mouse Models 7 http://dx.doi.org/10.5772/66438

Q72 have displayed somatic mosaicism that is tissue-specific, age-dependent and CAG repeat length dependent [25, 36, 42, 43]. Some neurological and motor impairment have also been described in these mice and might be correlated to the somatic mosaicism level [25, 36, 44]. A different HdhQ80 mouse model has been created by replacement of the murin exon 1 with the human exon 1 carrying ~80 CAG repeat using C57BL6/J mice. Small expansions upon paternal transmissions and CAG repeat contractions across maternal transmissions have been reported in about 20% of cases. As observed in HdhQ150, Hdh4/Q80 and Hdh6/Q72, HdhQ80 mice have shown somatic mosaicism that is age-dependent and biased toward expansions with the highest levels in the striatum and liver [35]. Compared to BACHD and knock-in mice described above, R6 and HdhQ111 mice are the most commonly used to identify the genetic modifiers of CAG repeat instability [24, 32, 45]. Therefore, we will review the somatic and intergenerational instability features for both these animal models in the next section.

2.1. R6 transgenic mouse lines

The first successful HD transgenic mouse model was created in 1996 and called R6 lines of HD transgenic mice [32]. These mice were obtained by random integration of a short 5' fragment of human *HTT* gene containing 1000 bp of 5'UTR, exon 1 with ~130 CAG repeat tracts and the beginning of intron 1 in a CBA/C57BL6 genetic background. Five lines of mice were obtained with different insertion sites and CAG repeat lengths. The R6/T line carries a truncated HD transgene without CAG repeat expansions, the R6/0 line carries 142 CAG repeats, R6/1 carries 116 repeats, the R6/2 carries 144 CAG and the R6/5 line carries multiple copies of transgene with 128, 132, 135, 137 and 156 CAG repeats, respectively. R6/0 mice have shown no transgene expression and no phenotype compared to R6/1, R6/2 and R6/5. These three mouse lines develop progressively neurological abnormalities and show a variable age of onset that depends on the CAG repeat length and on the transgene expression levels. R6/1 and R6/2 are the most studied of these lines to assess both HD pathogenesis and CAG repeat instability.

To evaluate intergenerational CAG repeat lengths, fluorescent PCR using DNA from tail biopsy at 3 weeks of age was performed in R6/0, R6/1, R6/2 parents and offspring. The comparison of CAG repeat lengths between parents and their progeny is limited in R6/5 mice due to the integration of multiple transgene copies in the genome of this line. Compared to R6/1 and R6/2, R6/0 mice do not show any evidence of CAG repeat instability and any transgene expression. As observed in HD patients, R6/1 and R6/2 mice mimic intergenerational instability biased toward expansions across paternal transmissions and toward contractions during R6/1 maternal transmissions (R6/2 female mice are infertile) with a mutation rate from 65 to 84% [45, 46]. Interestingly, the CAG repeat size changes depend on the gender of R6/1 embryos with a high expansion rate in males and high contraction rate in females from the same fathers suggesting that offspring sex-dependent genes modulate intergenerational instability in R6/1 mice [46]. In R6/2 mice, the size of transmitted CAG expansion increases with the age of transmitting males [45]. A selective R6/2 breeding enabled to obtain numerous R6/2 colonies with inheriting CAG repeat ranging from \sim 110 to 450 [47–49]. The size of CAG repeat is positively correlated with the severity of symptoms up to ~160 CAG repeats [47]. Surprisingly, some neurological symptoms and a lifespan are greatly ameliorated in R6/2 mice

8 Huntington's Disease - Molecular Pathogenesis and Current Models

carrying more than 200 CAG repeat expansions [47–49]. These unexpected results can be explained by transgene expression decrease observed in these mice [48]. A spontaneous contraction from 116 to ~89 CAG repeat was described in R6/1 mice [50]. These mice showed a decreased age of onset and a HD phenotypic improvement compared to R6/1 mice with 116 CAG repeat supporting the relationship between the CAG repeat size and the progression of symptoms.

Somatic instability of the CAG repeat tracts has been also reported in R6 lines carrying CAG repeat expansions excepted for the R6/0 line [45]. R6/1 and R6/2 recreated expansion-biased, age-dependent and tissue-specific somatic mosaicism as observed in HD patients [38, 51, 52]. Liver and striatum have shown the highest levels of instability biased toward expansions compared to other tissues that have shown low or no instability in both lines. Two distinct modes of somatic expansion have been described in tissues from R6/1 mice. Striatum and cortex have shown a periodic expansion, whereas the other tissues reproduce a short continuous expansion overtime suggesting different mechanisms of instability in these tissues [51]. Large spontaneous expansions (>200 CAG) have been described in striatum and cortex from R6/2 mice [52] consistent with the observations done in brain from HD patients [25, 43]. In R6/2 mice, the somatic mosaicism is correlated with the transmitted CAG repeat size but the somatic variation is not linear, particularly in striatum [52]. Interestingly, the frequency of CAG contractions increases in brain tissues and liver from mice with more than 500 CAG repeats [52] and could also explain the progressive reduction of neurological symptoms and prolonged lifespan in R6/2 mice with >200 CAG repeats [47–49]. Somatic instability has been noticed in dividing cells suggesting a role of DNA replication in the dynamic of triplet repeat instability. However, an increase of CAG repeat length has also been reported in terminally differentiated neurons from R6/1 mice suggesting the role of cellular processes independent of DNA replication in the somatic mosaicism [38]. Recently, an effect of mouse genetic backgrounds on the dynamics of CAG expansions has been reported in tissues from R6/1 mice with high CAG somatic mosaicism on a B6 background and low level in BALB/cBy backgrounds suggesting the existence of genetic modifiers of instability [53].

2.2. HdhQ92-111 mouse models

The first knock-in mice called HdhQ50 have been generated in 1997 using homologous recombination in ES cells to replace short murine CAG repeat by 48 CAG repeats in 129SvEv/CD1 mice [54]. In 1999, three other knock-in mouse models (HdhQ20, HdhQ92 and HdhQ111) using the same strategy have been generated with 18, 90 and 109 CAG repeat tracts, respectively [24]. These four knock-in mice share the identical murine genomic environment (91% of similarities with *HTT* human) and differ only by the size of CAG repeat length. Knock-in mice with >50 CAG repeats reproduce the pattern of intergenerational instability observed in HD patients. The mutation rate is only 4% in HdhQ50, 49% in HdhQ92 and 73% in the HdhQ111 supporting that intergenerational instability depends on CAG repeat length as described in R6 mice [24]. However, no age effect has been observed in these knock-in mice compared to R6/2 [24]. These divergent results could be explained by the CAG repeat genomic context and the genetic background. Interestingly, an effect of mouse genetic backgrounds on mutation

rate and range of CAG repeat length changes upon male transmission was reported in HdhQ111 supporting the role of genetic modifiers on CAG instability process [55].

Somatic CAG repeat variations have been observed in HdhQ92 and HdhQ111 mice in brain and some peripheral tissues with the highest accumulation of expansions in striatum and liver [24, 56, 57]. Both these tissues showed a bimodal distribution of repeat lengths compared to spleen and tail that showed a unimodal distribution [56]. CAG expanded alleles were broadly distributed in striatum compared to liver that showed distinct populations of CAG repeat expansions [56]. Somatic instability depends on the CAG repeat size and the age of animals and is tissue-specific as reported in R6 mice [24, 56, 57]. The relationship between somatic mosaicism and HD phenotype remains unclear but some data have reported that somatic mosaicism is not correlated with the initiation of disease but may be correlated with the progression of HD phenotypes [57, 58].

In conclusion, HD mouse models closely reproduced the dynamic of instability observed in HD patients. Intergenerational instability is biased toward expansions and depends on the CAG repeat length and the sex of transmitting parent. HD transgenic and knock-in mouse models also mimic the somatic instability of HD patients, with the highest somatic mosaicism in the striatum that is the most affected tissue in HD. Some differences in the dynamics of intergenerational instability between HD patients and HD mouse models can be noticed. Despite a high level of instability biased toward expansions in paternal transmissions and contractions in maternal transmissions in both species, the critical CAG repeat threshold length differs between human and mice corresponding to 35 CAG repeats in human and more than 80 CAG repeats in mice. Moreover, no spontaneous large CAG repeat expansion has been observed in HD mouse models during paternal transmissions, in contrast to HD patients. These differences may be explained by genetic and environmental factors. Despite these divergences, the development of HD mouse models provided a powerful tool to explore trinucleotide repeat dynamics. Several data have suggested that the size, sex and the age factors are not sufficient to explain the level of meiotic and mitotic instabilities observed in HD patients and mice supporting the contribution of genetic modifiers in CAG repeat instability processes. Among described mouse models, R6 and HdhQ111 were commonly used to investigate the role of genetic modifiers on the level of intergenerational and somatic instability in HD.

3. Genetic modifiers of CAG repeat instability

The absence of correlation between CAG repeat somatic mosaicisms and the corresponding tissue proliferative rates and the destabilization of CAG repeat in murine mature neurons support the involvement of DNA repair pathways in the CAG repeat instability processes (**Table 2**). To identify the DNA repair pathways involved in the germline and somatic CAG repeat instability, R6/1 or HdhQ111 mice were crossed with mouse lines deficient for individual DNA repair genes. CAG repeat length changes upon transmissions were determined by comparing the CAG repeat size in the HD transmitting mice with CAG repeat length in the HD progeny for each DNA repair genotype (+/+ to +/+ and -/- to -/- and/or +/- to +/-). Further-

more, different methods have been described to quantify the degree of somatic instability and have made it possible to compare the level of somatic mosaicism between HD mice mutated and not for DNA repair genes [53, 57, 59–61].

Gene modifiers	DNA repair	Gene status	Mouse models	Effect on CAG re	References	
	systems		_	Intergenerational instability	Somatic instability	_
Msh2	MMR	КО	Hdh ^{Q111}	↓ expansions ↑ contractions (male transmissions)	CAG repeat stabilization	[58, 67]
				No change (female transmissions)		
			R6/1	No expansion (male transmissions) ND (female transmissions)	↓ expansions	[59, 63]
Msh3			Hdh ^{Q111}	No significant change	CAG repeat	[64]
		C57BL/6J	R6/1	ND	stabilization	[53]
		BALB/cByJ				
Msh6		КО	Hdh ^{Q111}	No change (<i>Msh6-/-</i>)	No change	[64]
Mlh1				ND	↓ expansions	[69]
Mlh3						
Ogg1	BER		R6/1	No change	↓ expansions	[65, 70]
Neil1				↓ expansions (male transmissions)		[61]
Fen1				ND (female transmissions)	No change	[66]
Csb	NER			\uparrow expansions \downarrow contractions		[65]
Хрс				No change		[64]
ND, not det	ermined.					

Table 2. DNA repair genetic modifiers involved in CAG repeat instability in mouse models.

3.1. Genetic modifiers of intergenerational instability

Despite some controversial results, the analyses in *E. coli* and *S. cerevisiae* suggested an effect of mismatch repair (MMR) proteins on the dynamics of CAG repeat instability. MMR proteins preserve genome integrity by repairing erroneous insertion, deletion and misincorporation of bases that occur during replication and escape proofreading. Two MutS heterodimeres, MutS α (MSH2-MSH6) and MutS β (MSH2-MSH3) recognize replication errors and recruit MutL α (MLH1-PMS1/2) and MutL γ (MLH1-MLH3) to activate the repair pathway [62]. The breeding of HD mouse models in MMR-deficient genetic backgrounds has provided insight into the mechanisms of CAG repeat instability. In R6/1 mice, *Msh2* deficiency abolishes CAG Genetic Modifiers of CAG.CTG Repeat Instability in Huntington's Disease Mouse Models 11 http://dx.doi.org/10.5772/66438

repeat expansion in the male germline suggesting that MSH2 promotes CAG repeat expansion (no data for maternal transmissions) [59, 63]. A further study in HdhQ111 knock-in mice revealed that the effects of Msh2 mutation on the intergenerational dynamics seem to be more complex. The absence of two *Msh2* alleles suppresses the expansions in favor of contractions without changing the mutation rate (corresponding to expansion and contraction frequencies) in paternal transmissions [58, 64]. In contrast, a majority of contractions and a few expansions were detected in female germline transmissions in both Msh2+/+ and Msh2-/- backgrounds [24, 58]. Therefore, although MSH2 appears to be required in paternal CAG repeat expansions, the CAG repeat gains in female germline and CAG repeat contractions seem to be generated by Msh2-independent processes [58]. MSH2 binding partners, MSH6 or MSH3, did not alter the frequency of maternal changes, which is consistent with the lack of involvement of MSH2 in female germline. The effects of MSH3 and MSH6 on paternal transmissions of the expanded CAG repeat are more complex. The loss of *Msh6* or *Msh3* did not significantly affect the paternal mutation frequencies and the frequency of expansions and contractions. However, a shift from expansion to unchanged and contracted CAG repeat length is observed in Msh3-/- or Msh3+/transmissions compared to Msh3+/+ transmissions suggesting that some paternal expansions might depend on MSH3 protein. These results together suggest that the majority of paternal expansions occur via MSH2, independently of MSH3 and MSH6 partners in HdhQ111 mice and that other DNA repair proteins are involved in CAG repeat parental expansions and contractions observed in HD mice [64].

The involvement of base excision repair (BER) and nucleotide excision repair (NER) in CAG repeat instability have been tested in R6/1 mice bred in a BER gene (Ogg1 or Neil1) and NER gene (Csb or Xpc) deficient backgrounds. The loss of 7,8 dihydroxy-8-oxoguanin-DNA-glycosylase, OGG1 did not affect the dynamic of instability in the germ cells [65]. However, NEIL1, another glycosylase of BER contributed to paternal expansions in R6/1 mice. In the absence of *Neil1*, the CAG repeat tracts were more stable with a tendency toward contraction in male germline compared to *Neil1++* [61]. Interestingly, an increase of CAG repeat expansions and a decrease of contractions in paternal transmissions have been observed in *Csb*-deficient mice suggesting that CSB promotes CAG repeat contractions during paternal transmissions just like MSH2 promotes expansions in HD mouse models [65]. In contrast to *Cbs* results, *Xpc* did not affect the dynamic of CAG repeat instability in R6/1 mice. It has also been reported that FEN1, an endonuclease involved in the DNA replication but also in BER intermediates, may stabilize CAG repeat in the *Fen1+/-* male germline by preventing deletions and modestly increasing expansions but the effect seems to be low [66].

In conclusion, these data have shown that MSH2 and NEIL1 proteins are involved in the formation of intergenerational repeat expansions in HD mouse models with the highest effect of MSH2, suggesting that these genes are genetic modifiers of intergenerational instability in HD. Moreover, the shift toward contractions observed in the absence of *Msh2* and *Neil1* reveals that the repeat could be processed through a distinct pathway leading to contractions via CSB or other DNA repair proteins.

3.2. Genetic modifiers of somatic mosaicism

The analysis of CAG repeat instability has revealed a relationship between the severity of HD phenotypes and the level of expansion-biased somatic mosaicism in patients and mice. Thus, HD mouse models in DNA repair deficient background have also been used to identify genetic modifiers of somatic instability. In R6/1 and HdhQ111 mice, *Msh2* deficiency was initially reported to stabilize CAG repeat expansion in somatic tissues supporting that MSH2 also drives instability toward expansions, like in germline cells [58, 59, 67].

Compared to the results obtained in Msh6-deficient mice, the loss of both Msh3 alleles stabilize CAG repeat tracts in somatic tissues suggesting that MSH3 acts as an enhancer of CAG expansions-biased somatic instability but not MSH6 [53, 64, 68]. Interestingly, the absence of one allele of Msh3 is sufficient to decrease the somatic mosaicism in the striatum in contrast to *Msh2* supporting the idea that MSH3 levels modulate the degree of somatic instability and the progression of HD disease [64]. Various degree of repeat instability in different HdhQ111 and R6/1 mouse strains harboring the identical CAG repeat length suggest the existence of other candidate factors as a source for strain-specific variation in CAG repeat pattern [53, 55]. Interestingly, CAG repeat somatic mosaicism has been associated with Msh3 polymorphisms and the level of MSH3 protein [53]. It has been reported that expansion changes were higher in striatum and liver from R6/1 mice carrying the homozygous B6 Msh3 gene on a CBy genetic background than mice carrying the homozygous CBy Msh3 gene on a B6 genetic background (mice obtained by selective breeding). The loss of one B6 *Msh3* allele in mice on a CBy genetic background was sufficient to decrease CAG repeat instability, consistent with the results obtained in Msh3-deficient mice [53]. Thus, naturally occurring MSH3 protein polymorphisms modify the dynamic of CAG repeat instability in mice and could modulate HD pathogenesis in humans. Together, these data have shown that MSH2 and MSH3 proteins are strongly required in the generation of somatic expansions.

To identify other genetic modifiers of CAG repeat instability, linkage analyses have been performed in different HdhQ111 strains that showed CAG repeat instability variation [69]. A single quantitative trait locus on chromosome 9 and particularly in MutL homolog *Mlh1* gene has been identified and associated with CAG repeat instability. Then, somatic instability has been quantified in B6 HDHQ111 mice in the absence of one or two *Mlh1* alleles. Although one functional *Mlh1* allele was still sufficient to generate high levels of repeat expansion, the loss of both *Mlh1* alleles abolished CAG repeat expansion in striatum suggesting that MLH1 was required in somatic expansion. A second MutL homolog has been shown to act as an enhancer of CAG repeat expansions. Indeed, expansion-biased somatic mosaicism is reduced in *Mlh3* heterozygous knockout mice and totally abolished in *Mlh3* homozygous knockout mice suggesting that MLH3 is a limiting factor on the process of expansion as reported for MSH3 protein [69].

Other DNA repair systems, such as BER and NER have also been investigated in R6/1 mice to understand the somatic expansion variation observed between and within tissues. A loss of *Ogg1* suppressed CAG somatic expansions in 70% of R6/1 mice. The same study has reported that OGG1 initiated age-dependent CAG repeat expansion mice, suggesting that age-dependent somatic expansion associated with HD occurs in the process of removing

oxidized base lesions [70]. Deletion of *Neil1* also reduced somatic expansions in male and female R6/1 mice with a higher effect in different brain regions from male mice [61]. In contrast to the results obtained in male germline, the absence of *Csb* and *Fen1* did not affect the dynamic of somatic instability in tissues suggesting that the role of *Csb* is specific of paternal contractions [65] and that *Fen1* partially contributes to CAG repeat expansion upon parental transmissions [66].

In conclusion, MSH2 and MSH3, partner proteins in the MutS β MMR complex and MutL γ (MLH1-MLH3) are essential to promote expansions in HD mouse models suggesting that MutS β and MutL γ promote CAG expansion via the mismatch repair machinery. Furthermore, CAG repeat expansion depends only partially on OGG1, NEIL1 and FEN1 proteins suggesting that other DNA repair pathways are involved in the process of instability. Some genetic modifiers such as Ogg1 and Fen1 impact CAG repeat instability in either somatic or germline tissues, but not in both supporting that CAG repeat instability involves different genetic players between tissues and may occur via different mechanisms. It has also shown that the degree of somatic mosaicism appears to be modulated by Msh3 and Mlh1 variants in B6 mice where CAG repeat expansion levels are the highest suggesting that somatic instability variation observed in HD patients could be explained by DNA repair gene and/or protein variants. Different expression levels of MSH3 and MLH3 have been identified in mouse strains that exhibit different expansion frequencies supporting that the level of DNA repair proteins might be correlated with the degree of CAG repeat instability. Other studies also support a role for the stochiometries of DNA repair proteins in CAG repeat instability [4, 64, 71–73]. Few data have reported the role of genetic factors in CAG repeat contractions mainly observed in HD maternal transmissions and only Csb has been reported to promote contractions in paternal transmissions. CSB protein could act on CAG repeat contraction via BER, NER, or chromatin maintenance/remodeling activity independently of MSH2 protein.

4. Are genetic modifiers a therapeutic target?

The identification of genetic modifiers of underlying CAG repeat instability is important to uncover novel therapeutic targets to slow down somatic instability and to decrease the intergenerational expansions in favor of CAG repeat contractions to prevent the disease. It has been reported that *Msh2* alleles delay the accumulation of mutant protein and destruction of mutant huntingtin in striatum and in specific neuron type from knock-in HdhQ111 mice [58, 67]. Moreover, MLH1 also contributes to nuclear huntingtin and HD inclusion phenotypes [69]. Both data suggest that MSH2 and MLH1 may enhance the HD pathogenic process by modulating the somatic mosaicism in cooperation with MSH3 and MLH3 via the mismatch repair pathway. Among MMR proteins, MSH3 and MLH3 are currently the most promising targets to decrease CAG repeat expansions, thus delaying pathogenic process, given their minor roles in the initiation of human cancer. To date, no drug has been identified to decrease the expression of MLH3 and MSH3 protein and then the somatic instability. NEIL1 and OGG1, two glycosylases of the BER pathway partially contribute to CAG repeat expansions suggesting that oxidative base damage is responsible of some CAG repeat expansions. Antioxidants may

14 Huntington's Disease - Molecular Pathogenesis and Current Models

then decrease the expansion process. Mollersen and colleagues have suggested that several antioxidants like anthocyanin decrease CAG repeat expansion in the brain from R6/1 male mice [61]. The identification of new genetic factors involved in the formation of CAG repeat contractions and a better understanding of expansion mechanisms are essential. Novel therapies based on activating the DNA repair pathways promoting contractions might be expected to have lower risk of side effects than therapies based on inhibiting the DNA repair pathways that promote expansions.

5. Conclusion

The data summarized in this chapter have shown that *cis*-elements such as DNA sequence and transcription level, mismatch repair, base excision repair and nucleotide excision repair proteins can modulate the pathogenic expansion-biased somatic mosaicism and/or intergenerational expansions contributing to the progression of HD phenotype. Natural polymorphisms in *Msh3* and *Mlh1* genes have been associated with the degree of somatic expansions in HD mice suggesting that MMR variants are involved in the somatic mosaicism variation observed in HD patients and may modulate the disease severity and age of onset. Despite a great advance on the understanding of instability, the process remains complex. Then, further studies will be needed to assess how the various DNA repair and replication proteins collaborate all together in germline and/or somatic tissues to mediate CAG repeat expansions. Moreover, future studies will be essential to identify new factors that promote contractions in the germline and in somatic tissues, to reverse the HD expansion and to stop the disease.

Acknowledgements

The authors thank Geneviève Gourdon, Mario Gomes-Pereira and Diana Dinca for helpful comments and discussions. The authors also thank Christopher Pearson group, HD/CAG.CTG repeat colleagues, Canadian Institutes of Health Research, Imagine institute, INSERM, the Association Française contre les Myopathes (AFM) and the université Paris Descartes.

Author details

Elodie Dandelot^{1,2} and Stéphanie Tomé^{1,2*}

*Address all correspondence to: stephanie.tome@inserm.fr

1 Inserm UMR 1163, CTG Repeat Instability and Myotonic Dystrophy, Paris, France

2 Paris Descartes – Sorbonne Paris Cité University, Imagine Institute, Paris, France

References

- [1] Pearson CE, Nichol Edamura K, Cleary JD (2005) Repeat instability: mechanisms of dynamic mutations. Nat Rev Genet 6: 729–742.
- [2] Orr HT, Zoghbi HY (2007) Trinucleotide repeat disorders. Annu Rev Neurosci 30: 575– 621.
- [3] Usdin K, House NC, Freudenreich CH (2015) Repeat instability during DNA repair: insights from model systems. Crit Rev Biochem Mol Biol 50: 142–167.
- [4] Dion V (2014) Tissue specificity in DNA repair: lessons from trinucleotide repeat instability. Trends Genet 30: 220–229.
- [5] Warby SC, Montpetit A, Hayden AR, Carroll JB, Butland SL, et al. (2009) CAG expansion in the Huntington disease gene is associated with a specific and targetable predisposing haplogroup. Am J Hum Genet 84: 351–366.
- [6] Rawlins MD, Wexler NS, Wexler AR, Tabrizi SJ, Douglas I, et al. (2016) The prevalence of Huntington's disease. Neuroepidemiology 46: 144–153.
- [7] Gonzalez-Alegre P, Afifi AK (2006) Clinical characteristics of childhood-onset (juvenile) Huntington disease: report of 12 patients and review of the literature. J Child Neurol 21: 223–229.
- [8] The Huntington's Disease Collaborative Research Group (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell 72: 971–983.
- [9] Kay C, Collins JA, Miedzybrodzka Z, Madore SJ, Gordon ES, et al. (2016) Huntington disease reduced penetrance alleles occur at high frequency in the general population. Neurology 87: 282–288.
- [10] Rubinsztein DC, Leggo J, Coles R, Almqvist E, Biancalana V, et al. (1996) Phenotypic characterization of individuals with 30-40 CAG repeats in the Huntington disease (HD) gene reveals HD cases with 36 repeats and apparently normal elderly individuals with 36-39 repeats. Am J Hum Genet 59: 16–22.
- [11] Andrew SE, Goldberg YP, Kremer B, Telenius H, Theilmann J, et al. (1993) The relationship between trinucleotide (CAG) repeat length and clinical features of Huntington's disease. Nat Genet 4: 398–403.
- [12] Snell RG, MacMillan JC, Cheadle JP, Fenton I, Lazarou LP, et al. (1993) Relationship between trinucleotide repeat expansion and phenotypic variation in Huntington's disease. Nat Genet 4: 393–397.
- [13] Kremer B, Almqvist E, Theilmann J, Spence N, Telenius H, et al. (1995) Sex-dependent mechanisms for expansions and contractions of the CAG repeat on affected Huntington disease chromosomes. Am J Hum Genet 57: 343–350.

- 16 Huntington's Disease Molecular Pathogenesis and Current Models
 - [14] Leeflang EP, Zhang L, Tavare S, Hubert R, Srinidhi J, et al. (1995) Single sperm analysis of the trinucleotide repeats in the Huntington's disease gene: quantification of the mutation frequency spectrum. Hum Mol Genet 4: 1519–1526.
 - [15] Zuhlke C, Riess O, Bockel B, Lange H, Thies U (1993) Mitotic stability and meiotic variability of the (CAG)n repeat in the Huntington disease gene. Hum Mol Genet 2: 2063–2067.
 - [16] Novelletto A, Persichetti F, Sabbadini G, Mandich P, Bellone E, et al. (1994) Analysis of the trinucleotide repeat expansion in Italian families affected with Huntington disease. Hum Mol Genet 3: 93–98.
 - [17] Wheeler VC, Persichetti F, McNeil SM, Mysore JS, Mysore SS, et al. (2007) Factors associated with HD CAG repeat instability in Huntington disease. J Med Genet 44: 695– 701.
 - [18] Aziz NA, van Belzen MJ, Coops ID, Belfroid RD, Roos RA (2011) Parent-of-origin differences of mutant HTT CAG repeat instability in Huntington's disease. Eur J Med Genet 54: e413–418.
 - [19] Ramos EM, Cerqueira J, Lemos C, Pinto-Basto J, Alonso I, et al. (2012) Intergenerational instability in Huntington disease: extreme repeat changes among 134 transmissions. Mov Disord 27: 583–585.
 - [20] Leeflang EP, Tavare S, Marjoram P, Neal CO, Srinidhi J, et al. (1999) Analysis of germline mutation spectra at the Huntington's disease locus supports a mitotic mutation mechanism. Hum Mol Genet 8: 173–183.
 - [21] Yoon SR, Dubeau L, de Young M, Wexler NS, Arnheim N (2003) Huntington disease expansion mutations in humans can occur before meiosis is completed. Proc Natl Acad Sci U S A 100: 8834–8838.
 - [22] Telenius H, Kremer B, Goldberg YP, Theilmann J andrew SE, et al. (1994) Somatic and gonadal mosaicism of the Huntington disease gene CAG repeat in brain and sperm. Nat Genet 6: 409–414.
 - [23] De Rooij KE, De Koning Gans PA, Roos RA, Van Ommen GJ, Den Dunnen JT (1995) Somatic expansion of the (CAG)n repeat in Huntington disease brains. Hum Genet 95: 270–274.
 - [24] Wheeler VC, Auerbach W, White JK, Srinidhi J, Auerbach A, et al. (1999) Lengthdependent gametic CAG repeat instability in the Huntington's disease knock-in mouse. Hum Mol Genet 8: 115–122.
 - [25] Kennedy L, Evans E, Chen CM, Craven L, Detloff PJ, et al. (2003) Dramatic tissuespecific mutation length increases are an early molecular event in Huntington disease pathogenesis. Hum Mol Genet 12: 3359–3367.

- [26] Monckton DG, Wong LJ, Ashizawa T, Caskey CT (1995) Somatic mosaicism, germline expansions, germline reversions and intergenerational reductions in myotonic dystrophy males: small pool PCR analyses. Hum Mol Genet 4: 1–8.
- [27] Veitch NJ, Ennis M, McAbney JP, Shelbourne PF, Monckton DG (2007) Inherited CAG.CTG allele length is a major modifier of somatic mutation length variability in Huntington disease. DNA Repair (Amst) 6: 789–796.
- [28] Swami M, Hendricks AE, Gillis T, Massood T, Mysore J, et al. (2009) Somatic expansion of the Huntington's disease CAG repeat in the brain is associated with an earlier age of disease onset. Hum Mol Genet 18: 3039–3047.
- [29] Benitez J, Robledo M, Ramos C, Ayuso C, Astarloa R, et al. (1995) Somatic stability in chorionic villi samples and other Huntington fetal tissues. Hum Genet 96: 229–232.
- [30] Pouladi MA, Morton AJ, Hayden MR (2013) Choosing an animal model for the study of Huntington's disease. Nat Rev Neurosci 14: 708–721.
- [31] Hickey MA, Chesselet MF (2003) The use of transgenic and knock-in mice to study Huntington's disease. Cytogenet Genome Res 100: 276–286.
- [32] Mangiarini L, Sathasivam K, Seller M, Cozens B, Harper A, et al. (1996) Exon 1 of the HD gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. Cell 87: 493–506.
- [33] Lin CH, Tallaksen-Greene S, Chien WM, Cearley JA, Jackson WS, et al. (2001) Neurological abnormalities in a knock-in mouse model of Huntington's disease. Hum Mol Genet 10: 137–144.
- [34] Gray M, Shirasaki DI, Cepeda C andre VM, Wilburn B, et al. (2008) Full-length human mutant huntingtin with a stable polyglutamine repeat can elicit progressive and selective neuropathogenesis in BACHD mice. J Neurosci 28: 6182–6195.
- [35] Ishiguro H, Yamada K, Sawada H, Nishii K, Ichino N, et al. (2001) Age-dependent and tissue-specific CAG repeat instability occurs in mouse knock-in for a mutant Hunting-ton's disease gene. J Neurosci Res 65: 289–297.
- [36] Shelbourne PF, Killeen N, Hevner RF, Johnston HM, Tecott L, et al. (1999) A Huntington's disease CAG expansion at the murine Hdh locus is unstable and associated with behavioural abnormalities in mice. Hum Mol Genet 8: 763–774.
- [37] Gomes-Pereira M, Bidichandani SI, Monckton DG (2004) Analysis of unstable triplet repeats using small-pool polymerase chain reaction. Methods Mol Biol 277: 61–76.
- [38] Gonitel R, Moffitt H, Sathasivam K, Woodman B, Detloff PJ, et al. (2008) DNA instability in postmitotic neurons. Proc Natl Acad Sci U S A 105: 3467–3472.
- [39] Woodman B, Butler R, Landles C, Lupton MK, Tse J, et al. (2007) The Hdh(Q150/Q150) knock-in mouse model of HD and the R6/2 exon 1 model develop comparable and widespread molecular phenotypes. Brain Res Bull 72: 83–97.
- 18 Huntington's Disease Molecular Pathogenesis and Current Models
 - [40] Moffitt H, McPhail GD, Woodman B, Hobbs C, Bates GP (2009) Formation of polyglutamine inclusions in a wide range of non-CNS tissues in the HdhQ150 knock-in mouse model of Huntington's disease. PLoS One 4: e8025.
 - [41] Jin J, Peng Q, Hou Z, Jiang M, Wang X, et al. (2015) Early white matter abnormalities, progressive brain pathology and motor deficits in a novel knock-in mouse model of Huntington's disease. Hum Mol Genet 24: 2508–2527.
 - [42] Kennedy L, Shelbourne PF (2000) Dramatic mutation instability in HD mouse striatum: does polyglutamine load contribute to cell-specific vulnerability in Huntington's disease? Hum Mol Genet 9: 2539–2544.
 - [43] Shelbourne PF, Keller-McGandy C, Bi WL, Yoon SR, Dubeau L, et al. (2007) Triplet repeat mutation length gains correlate with cell-type specific vulnerability in Huntington disease brain. Hum Mol Genet 16: 1133–1142.
 - [44] Usdin MT, Shelbourne PF, Myers RM, Madison DV (1999) Impaired synaptic plasticity in mice carrying the Huntington's disease mutation. Hum Mol Genet 8: 839–846.
 - [45] Mangiarini L, Sathasivam K, Mahal A, Mott R, Seller M, et al. (1997) Instability of highly expanded CAG repeats in mice transgenic for the Huntington's disease mutation. Nat Genet 15: 197–200.
 - [46] Kovtun IV, Therneau TM, McMurray CT (2000) Gender of the embryo contributes to CAG instability in transgenic mice containing a Huntington's disease gene. Hum Mol Genet 9: 2767–2775.
 - [47] Cummings DM, Alaghband Y, Hickey MA, Joshi PR, Hong SC, et al. (2012) A critical window of CAG repeat-length correlates with phenotype severity in the R6/2 mouse model of Huntington's disease. J Neurophysiol 107: 677–691.
 - [48] Dragatsis I, Goldowitz D, Del Mar N, Deng YP, Meade CA, et al. (2009) CAG repeat lengths> or =335 attenuate the phenotype in the R6/2 Huntington's disease transgenic mouse. Neurobiol Dis 33: 315–330.
 - [49] Morton AJ, Glynn D, Leavens W, Zheng Z, Faull RL, et al. (2009) Paradoxical delay in the onset of disease caused by super-long CAG repeat expansions in R6/2 mice. Neurobiol Dis 33: 331–341.
 - [50] Vatsavayai SC, Dallerac GM, Milnerwood AJ, Cummings DM, Rezaie P, et al. (2007) Progressive CAG expansion in the brain of a novel R6/1-89Q mouse model of Huntington's disease with delayed phenotypic onset. Brain Res Bull 72: 98–102.
 - [51] Mollersen L, Rowe AD, Larsen E, Rognes T, Klungland A (2010) Continuous and periodic expansion of CAG repeats in Huntington's disease R6/1 mice. PLoS Genet 6: e1001242.

- [52] Larson E, Fyfe I, Morton AJ, Monckton DG (2015) Age-, tissue- and length-dependent bidirectional somatic CAG*CTG repeat instability in an allelic series of R6/2 Huntington disease mice. Neurobiol Dis 76: 98–111.
- [53] Tome S, Manley K, Simard JP, Clark GW, Slean MM, et al. (2013) MSH3 polymorphisms and protein levels affect CAG repeat instability in Huntington's disease mice. PLoS Genet 9: e1003280.
- [54] White JK, Auerbach W, Duyao MP, Vonsattel JP, Gusella JF, et al. (1997) Huntingtin is required for neurogenesis and is not impaired by the Huntington's disease CAG expansion. Nat Genet 17: 404–410.
- [55] Lloret A, Dragileva E, Teed A, Espinola J, Fossale E, et al. (2006) Genetic background modifies nuclear mutant huntingtin accumulation and HD CAG repeat instability in Huntington's disease knock-in mice. Hum Mol Genet 15: 2015–2024.
- [56] Lee JM, Pinto RM, Gillis T, St Claire JC, Wheeler VC (2011) Quantification of agedependent somatic CAG repeat instability in Hdh CAG knock-in mice reveals different expansion dynamics in striatum and liver. PLoS One 6: e23647.
- [57] Lee JM, Zhang J, Su AI, Walker JR, Wiltshire T, et al. (2010) A novel approach to investigate tissue-specific trinucleotide repeat instability. BMC Syst Biol 4: 29.
- [58] Wheeler VC, Lebel LA, Vrbanac V, Teed A, te Riele H, et al. (2003) Mismatch repair gene Msh2 modifies the timing of early disease in Hdh(Q111) striatum. Hum Mol Genet 12: 273–281.
- [59] Manley K, Shirley TL, Flaherty L, Messer A (1999) Msh2 deficiency prevents in vivo somatic instability of the CAG repeat in Huntington disease transgenic mice. Nat Genet 23: 471–473.
- [60] Morales F, Couto JM, Higham CF, Hogg G, Cuenca P, et al. (2012) Somatic instability of the expanded CTG triplet repeat in myotonic dystrophy type 1 is a heritable quantitative trait and modifier of disease severity. Hum Mol Genet 16: 3558–67.
- [61] Mollersen L, Rowe AD, Illuzzi JL, Hildrestrand GA, Gerhold KJ, et al. (2012) Neil1 is a genetic modifier of somatic and germline CAG trinucleotide repeat instability in R6/1 mice. Hum Mol Genet, 21: 4939–4947.
- [62] Fukui K (2010) DNA mismatch repair in eukaryotes and bacteria. J Nucleic Acids : 260512.
- [63] Kovtun IV, McMurray CT (2001) Trinucleotide expansion in haploid germ cells by gap repair. Nat Genet 27: 407–411.
- [64] Dragileva E, Hendricks A, Teed A, Gillis T, Lopez ET, et al. (2009) Intergenerational and striatal CAG repeat instability in Huntington's disease knock-in mice involve different DNA repair genes. Neurobiol Dis 33: 37–47.

- [65] Kovtun IV, Johnson KO, McMurray CT (2011) Cockayne syndrome B protein antagonizes OGG1 in modulating CAG repeat length in vivo. Aging (Albany NY) 3: 509–514.
- [66] Spiro C, McMurray CT (2003) Nuclease-deficient FEN-1 blocks Rad51/BRCA1mediated repair and causes trinucleotide repeat instability. Mol Cell Biol 23: 6063–6074.
- [67] Kovalenko M, Dragileva E, St Claire J, Gillis T, Guide JR, et al. (2012) Msh2 acts in medium-spiny striatal neurons as an enhancer of CAG instability and mutant huntingtin phenotypes in Huntington's disease knock-in mice. PLoS One 7: e44273.
- [68] Owen BA, Yang Z, Lai M, Gajek M, Badger JD, 2nd, et al. (2005) (CAG)(n)-hairpin DNA binds to Msh2-Msh3 and changes properties of mismatch recognition. Nat Struct Mol Biol 12: 663–670.
- [69] Pinto RM, Dragileva E, Kirby A, Lloret A, Lopez E, et al. (2013) Mismatch repair genes Mlh1 and Mlh3 modify CAG instability in Huntington's disease mice: genome-wide and candidate approaches. PLoS Genet 9: e1003930.
- [70] Kovtun IV, Liu Y, Bjoras M, Klungland A, Wilson SH, et al. (2007) OGG1 initiates agedependent CAG trinucleotide expansion in somatic cells. Nature 447: 447–452.
- [71] Mason AG, Tome S, Simard JP, Libby RT, Bammler TK, et al. (2013) Expression levels of DNA replication and repair genes predict regional somatic repeat instability in the brain but are not altered by polyglutamine disease protein expression or age. Hum Mol Genet 23: 1606–18.
- [72] Goula AV, Stys A, Chan JP, Trottier Y, Festenstein R, et al. (2012) Transcription elongation and tissue-specific somatic CAG instability. PLoS Genet 8: e1003051.
- [73] Goula AV, Berquist BR, Wilson DM, 3rd, Wheeler VC, Trottier Y, et al. (2009) Stoichiometry of base excision repair proteins correlates with increased somatic CAG instability in striatum over cerebellum in Huntington's disease transgenic mice. PLoS Genet 5: e1000749.





Compte rendu d'activités complémentaires au Doctorat

« La différence entre l'Homme et l'animal, c'est l'utilisation de la chose. » Un Corbeau bien renseigné **Dandelot, E**. and G. Gourdon (2018). "The flash-small-pool PCR: how to transform blotting and numerous hybridization steps into a simple denatured PCR." Biotechniques 64(6): 262-265.

Tome, S., **E. Dandelot**, C. Dogan, A. Bertrand, D. Genevieve, Y. Pereon, M. Simon, J. P. Bonnefont, G. Bassez and G. Gourdon (2018). "Unusual association of a unique CAG interruption in 5' of DM1 CTG repeats with intergenerational contractions and low somatic mosaicism." Hum Mutat 39(7): 970-982.

Dandelot, E. and S. Tomé (2017). "Genetic Modifiers of CAG. CTG Repeat Instability in Huntington's Disease Mouse Models." Huntington's Disease: Molecular Pathogenesis and Current Models: 1.

Participation à congrès

Dandelot, E. S. Tomé, C. Dogan, Y. Péréon, P.Cintas, A. Cobo Esteban, P. Arnaud, JP. Bonnefont, G. Bassez and G. Gourdon. Atypical Myotonic Dystrophy type 1 families and CTG repeats interruptions. Fourth Congress of Young Researchers of Imagine Institute, Paris, France -24th May 2018. (prix du meilleur poster)

Dandelot E, Tomé S, Gourdon G. Analyze the role of interruptions in triplet nucleotide repeats diseases: cellular model strategy and Flash-Small-Pool-PCR. International Conference on Unstable Microsatellites and Human Disease. Capri, Italy 21rth -26th April 2018.

Dandelot, E. Tome S, Dogan C, Péréon Y, Cintas P, Cobo Esteban A, Arnaud P, Bonnefont JP, Bassez G and Gourdon G. "Atypical Myotonic Dystrophy type 1 families and CTG repeat contractions". Imagine Young Researcher Seminars, Paris, France, 2nd March 2017 (presentation orale)

Dandelot, E. Tome S, Dogan C, Péréon Y, Cintas P, Cobo Esteban A, Arnaud P,Bonnefont JP, Bassez G and Gourdon G. Atypical Myotonic Dystrophy type 1 families and CTG repeat contractions. 10th Annual Graduate Student Symposium, Cambridge, UK, 14th-15th July 2016

Dandelot, E. Tome S, Dogan C, Péréon Y, Cintas P, Cobo Esteban A, Arnaud P, Bonnefont JP, Bassez G and Gourdon G. Atypical Myotonic Dystrophy type 1 families and CTG repeat contractions. Second Congress of Young Researchers of Imagine Institute, May 2016, Paris, France.

Dandelot, E. Tomé S, Dogan C, Péréon Y, Cintas P, Cobo Esteban A, Arnaud P, Bonnefont JP, Bassez G, Gourdon G, Atypical Myotonic Dystrophy type 1 families and CTG repeat contractions. . 5th International Congress of Myology, Lyon, France, 14th - 18th March 2016.

Formations Ecole Doctorale et Modules complémentaires

Disruptive Technologies and Public Policy – Certificat, SciencesPo, CRI, USPC, IIFR, 2017-2018

Médiation et journalisme – Certificat, CFDip-SPC , (du 11 au 13 Avril 2016 et du 23 au 25 Mai 2016

Activités sociétales et vulgarisation: produits destinés au grand public

Dandelot, E. Co-développement service médiation pour l'Institut Imagine, Paris, France **2019-2020**

Dandelot, E. Médiation Scientifique pour groupes scolaires, CRI, Paris, France 2018-2019

Dandelot, E. Médiation sur la Génétique, Journées des Maladies Rares, Institut Imagine, Paris, France **2019**

Dandelot, E. *« Tu seras microbe »* Médiation en immunologie par le jeu – Fête le Savoir/ Institut Pasteur, Vincennes, France **2018**

Dandelot, E. Médiation sur la Génétique, Institut Imagine, Journées du Patrimoine 2018

Dandelot, E. Médiation sur la Génétique, Palais de la Découverte/ Universcience. 2017-2018

Dandelot, E. 1000 Chercheurs dans les Écoles, AFM-Téléthon. Novembre 2017

Articles grand public

 Dandelot, E.
 "Six ans et demi, déjà chercheurs !," The Conversation, 27-Aug-2019. [Online].

 [Accessed:
 13-Sep-2019]
 <u>https://theconversation.com/six-ans-et-demi-deja-chercheurs-</u>

 108444
 108444

Dandelot, E. "Recherche en génétique humaine, techniques de pointe et parcours complexe » Découverte-Revue du Palais de la Découverte. [A paraître Décembre 2019].