

Détection de répétitions CGG par stratégie CRISPR-Cas9 sur séquenceur nanopore dans le cadre du syndrome de l'X fragile

P. Pugnère¹, J. Audoux², J.M. Holder², S. Beaumeunier², N. Philippe², P. Lochu¹, S. Kemeny¹, G. Egea¹

1) Département de Génétique, Imagenome, Gen-Bio, Clermont-Ferrand, France.
 2) SeqOne, IRMB, CHRU de Montpellier – Hôpital St Eloi, Montpellier, France.

INTRODUCTION

Le syndrome de l'X fragile est une maladie génétique rare caractérisée par une répétition anormale de triplets CGG au niveau de la région 5' non traduite du gène *FMR1* situé sur le chromosome X. Cette mutation dynamique peut évoluer d'un état de pré-mutation (55 à 200 répétitions) à un état de mutation complète de l'allèle (> 200 répétitions). Elle reste aujourd'hui la cause la plus fréquente de déficience intellectuelle héréditaire liée à l'X. Le dénombrement des triplets et l'évaluation du statut de méthylation des allèles sont les deux paramètres essentiels, obtenus séparément, pour l'établissement du diagnostic de la maladie. Le projet a donc pour but de développer un technique alternative plus rapide et moins coûteuse basée sur une approche ciblée par stratégie CRISPR-Cas9 et couplée au séquençage sur nanopores (*Oxford Nanopore Technologies*) pour la détermination simultanée du dénombrement de triplets CGG et du statut de méthylation de l'ADN.

MATÉRIELS & MÉTHODES

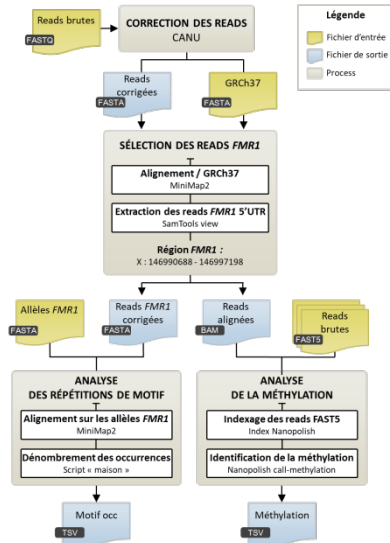
Préparation des bibliothèques

- ADN génomiques de haut poids moléculaire extraits à partir d'échantillons de sang total prélevés sur tube EDTA (Kit *Gentra® Puregene* - Qiagen).
- Région 5'UTR du gène *FMR1* séquencée pour des ADN de patients non mutés et des ADN contrôles pré-mutés et mutés (*Coriell Institute for Medical Research*) sans kit de *barcoding*. (nombre de répétitions CGG préalablement déterminé par techniques de PCR)
- Régions d'intérêt spécifiquement ciblées et clivées par CRISPR-Cas9 à l'aide du protocole d'enrichissement sans PCR médié par l'endonucléase Cas9 (ARNcr dessinés par *Oxford Nanopore Technologies* (ONT) et synthétisés par IDT).
- Séquençage nanopore effectué sur MinION (ONT).

Analyse bio-informatique des données de séquençage

Une approche spécifique a été développée pour dénombrer les triplets CGG et AGG au niveau de la région d'intérêt :

- Correction des *reads* réalisée par *CANU* ⁽¹⁾.
- Alignement réalisé sur le génome de référence humain (GRCh37) par *MiniMap2* ⁽²⁾.
- Utilisation de constructions d'allèles artificiels *FMR1* de différentes longueurs de répétitions (30, 80 et 200) pour le comptage des triplets.
- Après sélection du meilleur alignement, calcul du nombre d'occurrences comme : $ref_nb_occ + \frac{indel_dist}{3} \times (ref_nb_occ)$; *ref_nb_occ* : nombre d'occurrences de triplets dans l'allèle artificiel ; *indel_dist* : somme des bases insérées moins le nombre de bases supprimées dans l'alignement)
- Analyse de méthylation réalisée par *Nanopolish* ⁽³⁾.



Bibliographie

1. Koren S. et al. *Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation*. *Genome Research* (2017)
2. Li, H. *Minimap2: pairwise alignment for nucleotide sequences*. *Bioinformatics* (2018)
3. Loman JN. et al. *A complete bacterial genome assembled de novo using only nanopore sequencing data*. *Nature methods* (2015)

RÉSULTATS

Séquençage de 4 échantillons non mutés
 23, 28, 32, 53 répétitions de triplets CGG

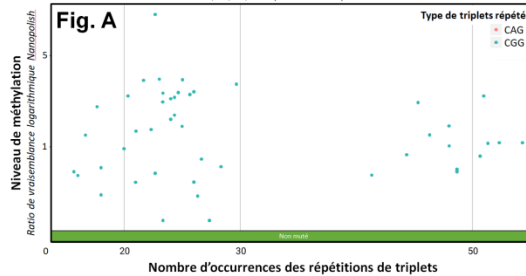


Fig. A: Les échantillons non mutés présentent tous des répétitions CGG mais ne peuvent être distingués les uns des autres qu'à partir de 20 répétitions. Le niveau de méthylation reste faible (ratio de vraisemblance logarithmique < 5), y compris pour des échantillons présentant un nombre de répétitions CGG appartenant à la zone grise (de 45 à 54 occurrences).

Séquençage de 2 échantillons mutés
 1 éch. pré-muté (30 et 73 répétitions de triplets CGG) + 1 éch. muté (> 270 répétitions de triplets CGG)

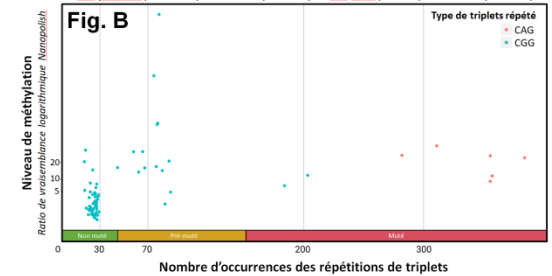


Fig. B: Les 3 types d'allèles sont identifiés par la méthode (30, 73 et > 200 répétitions de triplets). Les niveaux de méthylation les plus élevés sont observés pour les allèles pré-mutés et mutés (73 répétitions de triplets ou plus; ratio de vraisemblance logarithmique ≥ 10). Les répétitions de triplets CAG ne sont observées que pour les allèles présentant plus de 250 répétitions.

DISCUSSION

La technologie d'Oxford Nanopore a permis d'importantes économies de temps et d'argent dans le séquençage ciblé grâce à la stratégie CRISPR-Cas9 ⁽⁴⁾. Nos résultats montrent qu'il est possible de compter simultanément les expansions de répétitions de triplets CGG et d'évaluer l'état de méthylation des allèles pré-mutés et complètement mutés sans avoir besoin d'étapes particulières lors de la préparation des bibliothèques. Des applications cliniques plus étendues de cette méthode peuvent donc être envisagées dans le cadre de maladies impliquant l'expansion de triplets répétés de l'ADN, comme dans le cas de la maladie de *Huntington*. Cependant, le post-traitement bio-informatique des données avec correction de l'identification des bases reste essentiel en raison d'une mauvaise qualité du séquençage (Q score < 11). Par conséquent, il n'a pas été possible d'obtenir un nombre précis d'interruptions de triplets AGG lors de cette étude. D'autre part, il est à noter que le séquençage des plus petits fragments pourrait se faire préférentiellement au détriment des plus longs (Fig. B).

CONCLUSION

Cette étude démontre que le séquençage nanopore par stratégie CRISPR-Cas9 constitue un outil de diagnostic rapide et complet permettant la détection simultanée de mutations dynamiques ainsi que l'évaluation de l'état de méthylation des allèles complètement mutés. L'étude doit être com-

plétée par une optimisation de l'étape d'extraction de l'ADN et une évaluation précise de la sensibilité analytique. Cette approche pourrait être généralisée pour la détection de nombreuses autres mutations dynamiques impliquées dans diverses pathologies humaines.