

THESE DE DOCTORAT DE

NANTES UNIVERSITE

ECOLE DOCTORALE N° 605

Biologie-Santé

Spécialité : Biologie Cellulaire, Biologie du Développement

Par

Manon BAUDIC

Identification et caractérisation fonctionnelle de régions non-codantes à l'origine de cardiopathies rythmiques

Thèse présentée et soutenue à Nantes, le 13 mars 2023

Unité de recherche : L'institut du Thorax (INSERM UMR 1087 / CNRS UMR 6291)

Rapporteurs avant soutenance :

Francesca Rochais Chargée de recherche, INSERM, Marseille Medical Genetics

Guillaume Andrey Professeur assistant, Université de Genève

Composition du Jury :

Présidente : Catherine Boileau Professeure des universités- praticien hospitalier,
Hôpital Bichat-Claude Bernard, Université Paris Cité

Examinatrice : Estelle Gandjbakhch Professeure des universités- praticien hospitalier,
Hopital Pitié-Salpêtrière, Sorbonne Universités Paris

Dir. de thèse : Jean-Jacques Schott Directeur de recherche, INSERM, L'institut du thorax

Co-encadrant de thèse : Julien Barc Chargé de recherche, INSERM, L'institut du thorax

Titre : Identification et caractérisation fonctionnelle de régions non-codantes à l'origine de cardiopathies rythmiques

Mots clés : Trouble électrique cardiaque, variant non-codant, nouveau mécanisme moléculaire, TAD, épigénétique, région régulatrice, cellules souches pluripotentes induites humaines (hiPSC-CM)

Environ 10% des morts subites cardiaques (MSC) surviennent chez des individus sans cardiopathies ischémiques ni anomalies structurelles et sont considérées comme des pathologies électriques primaires (PEP), souvent héréditaires. L'origine génétique de ses pathologies a d'abord été associée à des mutations rares à effet fort dans les régions codantes du génome. Toutefois, certaines formes familiales montrent une expressivité variable de la pathologie et une pénétrance incomplète des variants rares suggérant l'implication de modulateurs. Cependant, la majorité des patients présentant des PEP reste à ce jour sans diagnostic moléculaire. L'une des hypothèses actuelles repose sur l'existence de mutations localisées dans les régions non-codantes du génome.

L'application de nouvelles technologies de séquençage de génomes entiers permet aujourd'hui d'aborder la génétique de ces patients en impasse de diagnostic moléculaire et le nouveau défi réside désormais dans l'annotation et la compréhension des mécanismes moléculaires de ces régions non-codantes à l'origine de cardiopathies.

Ainsi, à partir de 7 familles présentant une cardiopathie complexe associant des troubles électriques et des anomalies du développement, nous avons, identifié des délétions rares

chevauchantes (15-330kb) localisées dans un désert génique en 4q25. Au sein de ce projet international, je me suis attaché à caractériser le mécanisme épigénétique et les conséquences moléculaires sous-jacentes à ces délétions. Ce travail a conduit à étudier un modèle murin ainsi qu'un modèle isogénique hiPSC-CM. Grâce à ces modèles une dérégulation de l'expression du gène *PITX2* a été mis en évidence, causé par la délétion d'un site de liaison CTCF à l'origine d'un remodelage 3D de la chromatine.

En parallèle, j'ai appliqué mon expertise d'annotation génomique fonctionnelle des régions non-codantes au syndrome de Brugada (SBr), pour laquelle nous avons réalisé une étude d'association sur 2820 cas. Ce travail a permis l'identification de 21 variants fréquents-loci majoritairement non-codants. L'annotation génomique fonctionnelle de ces régions régulatrices montre leurs interactions avec des gènes cibles et, pour certains une régulation de leur expression. Le séquençage de génomes entiers de 354 cas index m'a également permis d'identifier une délétion rare d'une région régulatrice à l'origine du SBr.

En conclusion, ces travaux associant des approches génétiques, épigénétiques et fonctionnelles précisent l'architecture génétique complexe des cardiopathies rythmiques et ouvrent la voie pour une meilleure stratification du risque de la MSC.

Title: Identification and characterization of non-coding regions at the origin of rhythmic heart disease

Keywords: Cardiac electrical disorder, non-coding variant, new molecular mechanism, TAD, epigenetic, regulatory region, human induced pluripotent stem cells (hiPSC-CM)

About 10% of sudden cardiac deaths (SCD) occur in individuals without ischemic heart disease or structural abnormalities and are considered as primary electrical pathologies, that are often inherited. The genetic origin of electrical pathologies was initially associated with rare mutations with strong effect in the coding regions of the genome. However, some familial studies show variable expressiveness of the pathology and incomplete penetrance of rare variants, suggesting the involvement of modulators. However, the majority of patients with hereditary primary electrical diseases still remain without a molecular diagnosis. One of the current hypotheses is based on the existence of localized mutations in the non-coding regions of the genome.

The application of new sequencing technology allowed us to identify rare variants in these non-coding regions in patient genomes. A new challenge now lies in the annotation and understanding of the molecular mechanisms of these regions at the origin of heart disease.

Thus, from 7 families presenting with a complex heart disease associated electrical disorders to developmental anomalies, we identified rare overlapping deletions (15-330kb) located in a gene desert in the 4q25 region. In this international

collaboration, I focused on characterizing the epigenetic mechanism and molecular consequences underlying these deletions. This work led to the study of a murine model as well as an isogenic hiPSC-CM model. Thanks to these models, a dysregulation of the expression of the *PITX2* gene was highlighted, caused by the deletion of a CTCF binding site at the origin of a 3D remodeling of the chromatin.

In parallel, I applied my expertise in functional genomic annotation of non-coding regions to Brugada syndrome (BrS), for which we carried out an association study with 2820 cases. This work allowed the identification of 19 frequent non-coding variants-loci associated with BrS. The functional genomic annotation shows their interactions with target genes and for some, a regulation of their expression. Whole genome sequencing of 354 index cases also allowed me to identify a rare deletion of a regulatory region at the origin of BrS

In conclusion, this work combining genetic, epigenetic and functional approaches clarifies the complex genetic architecture of rhythmic heart disease and opens the way for a better risk stratification of SCD.