

# THESE DE DOCTORAT DE

L'UNIVERSITE  
DE BRETAGNE OCCIDENTALE  
COMUE UNIVERSITE BRETAGNE LOIRE

ECOLE DOCTORALE N° 605

*Biologie Santé*

Spécialité : Génétique, Génomique, Bioinformatique

Par

**Tania CUPPENS**

## **Conséquences du contexte haplotypique sur la fonction des protéines : application à la mucoviscidose**

Thèse présentée et soutenue à Brest, le 7 Mai 2019

Unité de recherche : INSERM U1078, équipe Génétique Moléculaire et Epidémiologie Génétique

### **Rapporteurs avant soutenance :**

Pascale FANEN      Professeure - Praticienne hospitalière, Université Paris-Est, Inserm U955  
Pascal RIHET      Professeur, Aix-Marseille Université, UMR1090

### **Composition du Jury :**

Président :

Examineurs :

Isabelle CALLEBAUT      Directrice de recherche, Sorbonne Université, UMR7590

Pascale FANEN      Professeure - Praticienne hospitalière, Université Paris-Est, U955

Claude FEREC      Professeur - Praticien hospitalier, Université Bretagne Occidentale, U1078

Jean MULLER      Maître de conférences - Praticien hospitalier, Université de Strasbourg, UMR112

Pascal RIHET      Professeur, Aix-Marseille Université, UMR1090

Dir. de thèse : Emmanuelle GENIN  
Co-dir. de thèse : Pascal TROUVE

Directrice de recherche, Université Bretagne Occidentale, U1078  
Ingénieur de recherche, Université Bretagne Occidentale, U1078

**Titre :** Conséquences du contexte haplotypique sur la fonction des protéines : application à la mucoviscidose

**Mots clés :** Outil bioinformatique, Contexte haplotypique, Visualisation, Mucoviscidose, Fonction protéique

**Résumé :** Notre génome contient des centaines de milliers de variants génétiques, qui pour la plupart, n'ont aucun impact sur notre santé. Après séquençage, il faut les filtrer pour ne conserver que ceux qui sont potentiellement impliqués dans une maladie. On utilise des annotateurs qui prédisent l'impact des variants. Ces prédictions sont faites sans tenir compte des variants en cis dans le même gène. Pourtant, des variants neutres peuvent, lorsqu'ils sont réunis chez un individu, devenir délétères. J'ai donc développé l'outil bioinformatique GEMPROT qui permet de visualiser l'effet des variants génétiques sur la séquence protéique et de mettre en évidence les combinaisons de variants touchant un même domaine fonctionnel.

J'ai ensuite étudié l'impact de deux variants associés à la p.Phe508del (508del) sur la protéine CFTR.

Le variant p.Val470M est présent sur tous les haplotypes portant la délétion mais pas sur la séquence de référence, qui est généralement utilisée pour la construction de plasmides. Nous avons montré des différences de fonction de la protéine CFTR selon l'acide aminé en position 470. La fonction est augmentée avec une Valine et il convient donc de s'assurer, lors de la construction de plasmides, que le contexte haplotypique des variants étudiés est bien respecté. Le variant p.Ile1027Thr conduit à une dégradation de la fonction de la protéine 508del. Ce variant n'est présent que sur une partie des haplotypes 508del et pourrait donc avoir un effet modificateur de l'expression de la délétion. En conclusion, nous montrons l'importance de la prise en compte des contextes haplotypiques dans l'étude des maladies et proposons un outil bioinformatique pour le faire.

**Title :** Consequences of the haplotype context on protein function: application to cystic fibrosis

**Keywords :** Bioinformatic tool, Haplotype context, Visualization, Cystic Fibrosis, Protein function

**Abstract :** We all carry hundreds of thousands genetic variations in our genome that, for the most of them, have no impact on our health. After sequencing, they must be filtered to only retain those potentially involved in a disease. We use annotators that predict the impact of variants. These predictions are done for each variant taken independently without considering cis variants in the same gene. However, neutral variants can become deleterious when associated together. I have developed the bioinformatics tool GEMPROT, which makes it possible to visualize the effect of genetic variants on the protein sequence and to highlight combinations of variants affecting the same functional domain. so.

I then studied the impact of two variants associated with p.Phe508del (508del) on CFTR protein function.

The variant p.Val470M is present on all carrying deletion haplotypes but not on the reference sequence, which is generally used for the construction of plasmids. We have shown differences in the function of the mutated CFTR protein 508del according to the amino acid at position 470. The function is increased with a Valine and it is therefore necessary to ensure, when constructing plasmids, that the haplotype context of the studied variants is well respected. The variant p.Ile1027Thr leads to a degradation of the function of the 508del protein. This variant is present only on a portion of the 508del haplotypes and could therefore have a modifying effect on deletion expression. In conclusion, we show the importance of considering haplotype contexts in the diseases studies and propose a bioinformatics tool to do