



Etude informatique des effets cliniques et omiques des modulateurs de CFTR dans la mucoviscidose et recherche de nouvelles cibles

Matthieu Cornet

► To cite this version:

Matthieu Cornet. Etude informatique des effets cliniques et omiques des modulateurs de CFTR dans la mucoviscidose et recherche de nouvelles cibles. Bio-informatique [q-bio.QM]. Université Paris sciences et lettres, 2022. Français. NNT : 2022UPSLM069 . tel-04053509

HAL Id: tel-04053509

<https://pastel.hal.science/tel-04053509>

Submitted on 31 Mar 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.



THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à MINES ParisTech

**Etude informatique des effets cliniques et omiques des
modulateurs de CFTR dans la mucoviscidose et
recherche de nouvelles cibles**

Soutenue par

Matthieu CORNET

Le 13 décembre 2022

Ecole doctorale n° 621

ISMME

Spécialité

Bio-informatique

Composition du jury :

Zhou, XU

Directeur de Recherche, CNRS *Rapporteur, Président*

Eric, BONNET

Directeur de Recherche, CEA *Rapporteur*

Philippe, REIX

Professeur, Hospices Civiles Lyon *Examineur*

Isabelle, SERMET-GAUDELUS

MD, Professeure, Institut Necker Enfants Malades
Co-Directrice de thèse

Véronique, STOVEN

Professeure, Mines Paris PSL *Directrice de thèse*

*A mes parents et à ma sœur,
Pour votre soutien inconditionnel*

Résumé

La mucoviscidose est une maladie autosomique récessive mortelle provoquée par une mutation du gène Cystic Fibrosis Transmembrane Regulator (CFTR), codant pour la protéine du même nom. Cette dernière est un transporteur transmembranaire d'ions Cl⁻. La plus fréquente de ces mutations est la p.Phe508delCFTR (F508del). L'absence ou le dysfonctionnement de CFTR sont associés à une inflammation et des infections chroniques, ainsi qu'à une bronchopathie chronique obstructive, entraînant des lésions au niveau des poumons et un déclin des fonctions respiratoires.

Le pronostic des patients porteurs de la mutation F508del a récemment été grandement amélioré grâce à la mise au point de thérapies basées sur de petites molécules appelées modulateurs de CFTR. Cependant, la réponse à ces traitements n'est pas homogène parmi les patients, et leur mécanisme d'action n'est pas entièrement compris. En particulier, nous ne disposons pas de marqueur biologique ou clinique qui fasse consensus dont la variation permettrait de suivre la réponse globale d'un patient aux traitements par modulateur, ou d'indiquer précocement le niveau de réponse du patient à ces traitements. L'amélioration des fonctions respiratoires, qui était jusqu'à présent prise comme référence, atteint ses limites lorsque l'on considère de jeunes patients dont les fonctions respiratoires sont encore bonnes : il est alors difficile de mesurer une amélioration, et on ne peut alors que surveiller la vitesse de déclin de ces fonctions. Ceci n'est pas satisfaisant car on souhaiterait évaluer la qualité de la réponse du patient à relativement court terme (en mois), pour éviter des traitements inutiles et proposer éventuellement des alternatives. Les limites sont par ailleurs évidemment atteintes lorsqu'il s'agit de très jeunes patients qui ne peuvent pas subir d'épreuves respiratoires. Par ailleurs, il est important de caractériser d'éventuels bénéfices cliniques ou biologiques autres, et d'étudier s'ils sont liés ou non à la réponse respiratoire. L'identification de tels bénéfices permettrait éventuellement de définir de nouveaux marqueurs de la réponse au traitement. L'objectif de cette thèse a été de caractériser les réponses cliniques aux modulateurs, en particulier chez les jeunes patients, ainsi que de mieux comprendre leurs mécanismes d'action au niveau cellulaire.

Je me suis dans un premier temps intéressé à la caractérisation de la réponse clinique des patients traités par une combinaison de ces modulateurs, dans l'espoir de caractériser différents profils de réponse clinique. Nous avons pu mettre en évidence plusieurs profils, dont certains indépendants des fonctions respiratoires. J'ai ensuite recherché de potentiels nouveaux

biomarqueurs de réponse dans la sueur, car CFTR étant exprimé dans les glandes sudoripares, leur fonctionnement est perturbé dans la mucoviscidose, et il était donc plausible que l'effet des médicaments y soit détectable précocement.

Pour mieux comprendre le fonctionnement des modulateurs, j'ai réalisé une étude protéomique et transcriptomique sur des cellules épithéliales nasales de 10 patients porteurs de la mutation F508del, traitées ou non avec les modulateurs. En particulier, l'activité des mitochondries est apparue fortement modifiée après traitement, ce qui n'était pas connu. Nous avons donc pu mettre en évidence l'effet du traitement sur les mitochondries, et l'importance de ces dernières dans la correction des défauts induits par le dysfonctionnement de CFTR dans la mucoviscidose.

Abstract

Cystic Fibrosis is a fatal autosomal recessive disease caused by a mutation in the Cystic Fibrosis Transmembrane Regulator (CFTR) gene, coding for the CFTR protein. CFTR is a transmembrane transporter of Cl⁻ ions. The most frequent of these mutations is p.Phe508delCFTR (F508del). Absence or dysfunction of CFTR is associated with chronic inflammation and infections, as well as chronic obstructive pulmonary disease, leading to lung damage and decline in lung function.

The prognosis of patients with the F508del mutation has recently been greatly improved by the development of therapies based on small molecules called CFTR modulators. However, the response to these therapies is not homogeneous among patients, and their mechanism of action is not fully understood. In particular, we do not have an universally accepted biological or clinical marker whose variation would allow us to monitor a patient's overall response to modulator treatments, or to give an early indication of the patient's level of response to these treatments. The improvement in respiratory function, which has been used as a reference until now, reaches its limits when considering young patients whose respiratory function is still good: it is then difficult to measure an improvement, and one can only monitor the rate of decline of these functions. This is not satisfactory because we would like to evaluate the quality of the patient's response in a relatively short term (in months), to avoid unnecessary treatments and possibly propose alternatives. The limits are obviously reached when it comes to very young patients who cannot undergo respiratory tests. Furthermore, it is important to characterize possible clinical or biological benefits, and to study whether or not they are related to the respiratory response. The identification of such benefits would eventually lead to the definition of new markers of response to treatment. The objective of this thesis was to characterize the clinical responses to modulators, particularly in young patients, and to better understand their mechanisms of action at the cellular level.

I was first interested in characterizing the clinical response of patients treated with a combination of these modulators, in the hope of characterizing different clinical response profiles. We were able to identify several profiles, some of them being independent of respiratory functions.

I then looked for potential new biomarkers of response in sweat: as CFTR is expressed in the sweat glands, whose function is disturbed in cystic fibrosis, it was plausible that the effect of drugs could be detected early in the sweat of patients.

To better understand the mechanism of action of the modulators, I performed a proteomic and transcriptomic study on nasal epithelial cells from 10 patients with the F508del mutation, treated or not with modulators. In particular, the activity of mitochondria appeared strongly modified after treatment, which was not known prior to this analysis. We were thus able to highlight the effect of the treatment on mitochondria, and the importance of the latter in the correction of the defects induced by the dysfunction of CFTR in cystic fibrosis.

Liste des publications

Matthieu Cornet, Geneviève Robin, Fabiana Ciciriello, Tiphaine Bihouee, Christophe Marguet, Valérie Roy, Muriel Lebourgeois, Frédérique Chedevergne, Anne Sophie Bonnel, Mairead Kelly, Philippe Reix, Vincenzina Lucidi, Véronique Stoven, Isabelle Sermet-Gaudelus

Profiling the response to lumacaftor-ivacaftor in children with cystic between fibrosis and new insight from a French-Italian real-life cohort

Pediatr Pulmonol 2022 Aug 22. doi: 10.1002/ppul.26123.

Matthieu Cornet, Thao Nguyen-Khoa, Mairead Kelly-Aubert, Vincent Jung, Frédérique Chedevergne, Muriel Le Bourgeois, Laura Aoust, Kévin Roger, Chiara Ida Guerrera and Isabelle Sermet-Gaudelus

Proteomic Profiling of Sweat in patients with Cystic Fibrosis provides new insights into epidermal homeostasis

Skin Health and Disease, 2022, doi: 10.1002/ski2.161

Cornet Matthieu*, Najm Matthieu*, Albergante Luca, Zinovyev Andrei, Sermet-Gaudelus, Isabelle, Stoven Véronique, Calzone Laurence, Martignetti Loredana

* co-first

Representation and quantification Of Module Activity from omics data with rROMA

Preprint 2022 Oct 25. Doi: 10.1101/2022.10.24.513448

Also With Editor, bioinformatics at the time of writing this manuscript

David Drummond, Jérémy Dana, Laureline Berteloot, Elena K Schneider-Futschik, Frédérique Chedevergne, Céline Bailly-Botuha, Thao Nguyen-Khoa, **Matthieu Cornet**, Muriel Le Bourgeois, Dominique Debray, Muriel Girard, Isabelle Sermet-Gaudelus

Lumacaftor-ivacaftor effects on cystic fibrosis-related liver involvement in adolescents with homozygous F508 del-CFTR

J Cyst Fibros 2022 Mar;21(2):212-219. doi: 10.1016/j.jcf.2021.07.018.

Remerciements

Cette section est particulière, étant donné le contexte de la fin de ma thèse. Mais en cela elle a aussi une signification spéciale à mes yeux.

Je souhaiterais tout d'abord remercier le jury. Un grand merci à Eric Bonnet d'avoir accepté d'être rapporteur, à Philippe Reix de s'être adapté aux contraintes de mon jury de thèse, et un grand merci à Zhou Xu de porter la double casquette rapporteur/président pour cette soutenance.

Véronique, je ne sais tout simplement pas comment je pourrais te remercier pour tout ce que tu as fait pour moi. C'est grâce à toi si j'ai eu l'opportunité de faire cette thèse, à partir d'un simple mail à tes anciens élèves proposant de s'intéresser à la mucoviscidose. Tu m'as accompagné dans les différentes étapes pour obtenir un financement. Tu as su me passionner pour la lutte contre cette maladie dont je ne connaissais pas grand chose avant de débiter cette thèse. Tu as toujours essayé d'accommoder le contenu de ma thèse pour qu'elle corresponde à mes besoins. Tu étais toujours là lorsque j'avais des questions. Même dans les moments difficiles pour moi, tu étais à l'écoute, toujours compréhensive. Tu étais encore derrière moi sur cette fin de thèse et tu m'as encouragé à aller jusqu'au bout. C'est simple : sans tes mots dans les derniers mois, ce manuscrit n'aurait jamais vu le jour. Pour tout ça je n'ai qu'un seul mot : Merci.

Isabelle, tu as été l'autre moteur principal de ma thèse. J'ai été et je serai toujours impressionné par ton niveau de maîtrise extraordinaire sur un sujet aussi complexe que la mucoviscidose, sous tous ses aspects. Tu trouvais toujours la bonne façon de m'aider pour que je ne me retrouve jamais bloqué en biologie. Tu m'as appris à écrire de véritables articles scientifiques, tu n'as pas hésité à passer des dizaines d'heures pour m'aider lors de leur rédaction. Tu avais toujours de nouvelles idées pertinentes pour faire avancer le projet. C'était véritablement un privilège de pouvoir travailler à tes côtés.

Merci Mairead pour toutes les connaissances que tu m'as apportées en biologie, et pour ton aide précieuse pour toute la partie biologie, en particulier pour la culture cellulaire. Mais surtout merci pour tous les bons moments passés au troisième étage, pour ces découvertes culinaires

extraordinaires, pour tes super anecdotes. J'espère qu'on continuera à pouvoir travailler ensemble de temps en temps !

Merci Matthieu pour tout ton travail sur la partie informatique. Nous formions un bon duo, la team Matthieu Matthieu ! Tes codes d'analyses omiques m'ont été d'une grande aide, et nos discussions sur les algorithmes étaient toujours très enrichissantes. Merci aussi d'avoir été là en dehors du travail. J'aurais aimé pouvoir jouer au tennis plus souvent avec toi, j'espère qu'on pourra remettre ça !

Merci à ma voisine de bureau, Agathe, toujours prête à m'aider, quel que soit le besoin. Nos discussions le soir avant de fermer le labo vont me manquer, nos sorties course à pied du vendredi aussi !

Merci à Emma et Charles, mes camarades de course du mercredi. C'était toujours une véritable bouffée d'air frais très appréciable en milieu de semaine !

Merci à Elise, Aurélie, Anita, Sophie, Apoline, Danielle, pour tous les bons moments passés au 3ème étages à Necker, et merci à tout le reste de l'équipe Sermet pour votre soutien et votre aide précieuse !

Merci aussi à Laurence et Loredana pour votre aide précieuse sur la biologie des système et sur rRoma. Vous étiez toujours de très bons conseils, et prêtes à m'aider à réorienter ma thèse dans une direction qui me conviendrait mieux. C'est en partie grâce à vous si j'ai pu aller au bout.

Merci à toute l'équipe du CBIO ! Je me souviendrai longtemps de cette journée à Fontainebleau à grimper sur des rochers avec toute l'équipe. J'ai senti que je rencontrais enfin tout le monde, et j'ai pu voir une équipe vraiment géniale !

Un grand merci à Geneviève pour ton énorme soutien au début de ma thèse ! Lorsque j'étais perdu et que je ne savais plus quoi faire à partir de cette maudite PCA qui me sortait par les yeux, tu étais là pour m'aider à y voir plus clair, pour me conforter dans mes choix ou au contraire pour me réorienter dans la bonne direction. Tu étais le phare qui m'était nécessaire pour retrouver mon chemin dans ce début de thèse tumultueux.

Mais surtout merci Morgane, merci papa, merci maman, merci d'être toujours là pour moi. Merci de toujours me soutenir, merci de me comprendre et d'accepter mes choix. Merci de me permettre de toujours avoir ce cocon, cet endroit où retourner, que je sois dans une bonne ou une mauvaise période, cet endroit hors du temps, où je sais que je me sentirai nécessairement bien. Finalement, ce n'est pas si grave que ça de perdre à Just Dance de temps en temps si cela me permet de passer plus de temps avec vous ! Si je suis là où je suis aujourd'hui, c'est surtout grâce à vous. Si j'ai la force d'écrire ce manuscrit aujourd'hui, c'est aussi grâce à vous. Je ne sais pas ce que le futur me réserve, et il me fait parfois peur, mais savoir que vous serez toujours derrière moi me permet de continuer à avancer.

Cette thèse a été menée avec le soutien de la Fondation pour la Recherche Médicale. Merci beaucoup à la Fondation de m'avoir accordé une bourse pour mener à bien mon doctorat.

Table des matières

Résumé	2
Abstract	4
Liste des publications	6
Remerciements	7
Table des matières	10
Liste des figures	14
Liste des annexes	16
Liste des abréviations	17
I. Introduction	19
1. La mucoviscidose	19
1.1 Population touchée	19
1.2 Diagnostic de la maladie	20
1.3 Symptômes	21
1.3.1 Symptômes pulmonaires et respiratoires	22
1.3.2 Symptômes pancréatiques	23
1.3.3 Symptômes intestinaux	23
1.3.4 Symptômes du foie	24
1.3.5 Symptômes génitaux	24
1.3.6 Symptômes de la peau	24
1.3.7 Autres Symptômes	24
2. CFTR	26
2.1 Le gène CFTR	26
2.2 La protéine CFTR	26
2.3 Biogénèse de CFTR	29
2.4 Les classes de mutation	31
2.4.1 Classe I : Mutation menant à l'absence de synthèse de la protéine malgré la production d'ARNm	31
2.4.2 Classe II : Mutation affectant la maturation de la protéine CFTR	31
2.4.3 Classe III : Mutation entraînant un défaut de la fonction canal chlorure de CFTR	32
2.4.4 Classe IV : Mutation limitant la conductance du canal chlorure de CFTR	32

2.4.5 Classe V : Mutation limitant la production de la protéine CFTR	32
2.4.6 Classe VI : Mutation menant à une protéine CFTR instable à la membrane	32
2.4.7 Classe VII : Mutation menant à l'absence de production d'ARNm	32
3. Traitement par modulateurs	34
3.1 Mesures en clinique pour évaluer l'effet des traitements	34
3.2 Historique du développement des traitements par modulateurs	39
3.2.1 Ivacaftor	39
3.2.2 Lumacaftor	42
3.2.3 Tézacaftor	44
3.2.4 Elexacaftor	45
3.3 L'essor des combothérapies	46
3.3.1 Lumacaftor/Ivacaftor (Orkambi)	46
3.3.2 Tezacaftor/Ivacaftor (Symkevi)	48
3.3.3 Triple combinaison VX-770/VX-661/VX-445 (TRIKAFTA)	49
3.3.4 Développement d'autres modulateurs	50
3.4 Nouveaux défis pour la détection de bénéfices cliniques	51
3.5 La recherche d'un score de réponse aux traitements	54
4. Approches omiques et intérêt dans la mucoviscidose	57
4.1 L'analyse protéomique	57
4.1.1 L'analyse Bottom-up	58
4.1.2 Protéomique dans la mucoviscidose	62
4.2 L'analyse transcriptomique	68
4.2.1 Réalisation pratique	68
4.2.2 Transcriptomique dans la mucoviscidose	69
5. Méthodes et algorithmes employés	73
5.1 L'Analyse en Composantes Principales	73
5.1.1 Explication pour les non mathématiciens	73
5.1.2 Détails mathématiques de l'ACP	77
5.2 L'algorithme de clustering k-means	78
5.3 Tests statistiques utilisés	78
5.3.1 Cas où l'hypothèse de normalité n'est pas vérifiée	79
5.3.2 Cas de tests multiples	79
II. Objectifs de la thèse	80

1. Peut-on caractériser l'hétérogénéité de la réponse clinique aux modulateurs de CFTR ?	80
2. Peut-on mieux comprendre le mécanisme d'action des modulateurs de CFTR à travers l'étude de la sueur ?	81
3. Peut-on mieux comprendre le mécanisme d'action des modulateurs de CFTR à travers l'étude de cellules épithéliales respiratoire ?	81
4. Quels outils utiliser lorsqu'une analyse gène par gène est insuffisante ?	82
III. Résultats	83
1. Analyse de la réponse clinique aux traitements par modulateurs de CFTR	83
1.1 Introduction	83
1.2 Matériels et méthodes	85
1.2.1 DESCRIPTION DES DONNEES	85
1.2.2 UTILISATION DE L'ACP	86
1.2.3 GROUPEMENT DES PATIENTS	87
1.2.4 ANALYSES STATISTIQUES	87
1.2.5 IMPUTATION DES VALEURS MANQUANTES	87
1.3 Résultats	90
1.3.1 CARACTERISTIQUES DE LA COHORTE	90
1.3.2 ANALYSE DES DONNEES DES PATIENTS AGES DE 12 A 18 ANS	91
1.3.3 ANALYSE DES DONNEES DES PATIENTS AGES DE 6 A 12 ANS	95
1.4 Discussion des résultats	96
1.5 Conclusion de l'étude	101
2. Analyse protéomique de la sueur de patients	102
2.1 Introduction	102
2.2 Matériels et méthodes	103
2.2.1 CARACTERISATION DES PATIENTS	103
2.2.2 PREPARATION DES ECHANTILLONS EN VUE D'UNE ANALYSE PAR SPECTROMETRIE DE MASSE	104
2.2.3 ANALYSE CHROMATOGRAPHIE/SPECTROMETRIE DE MASSE	105
2.2.4 ANALYSE DES DONNEES	105
2.3 Résultats	106
2.4 Discussion de l'étude	110

2.5 Cas de l'analyse de la trithérapie	113
2.6 Conclusion	118
3. Analyse multi-omique de cellules primaire de patients, évaluation des effets de traitements par modulateurs	119
3.1 Introduction :	119
3.2 Matériels et méthodes	120
3.2.1 PATIENTS DE L'ETUDE	120
3.2.2 CULTURE CELLULAIRE	121
3.2.3 ANALYSE RNA-seq	125
3.2.4 ANALYSE PROTEOMIQUE	126
3.2.5 ANALYSE INFORMATIQUE	127
3.3 Résultats	128
3.4 Discussion	130
3.5 Conclusion	133
4. rRoma, un algorithme pour l'étude de groupement des gènes lors d'analyses omiques	135
4.1 Introduction	135
4.2 Méthodes	138
4.2.1 UTILISATION DE L'ACP	138
4.2.2 PRETRAITEMENT DES DONNEES POUR RROMA	140
4.2.3 ORIENTATION DES COMPOSANTES	140
4.2.4 RETRAIT DES ECHANTILLONS « ANORMAUX »	141
4.2.5 SUPPRESSION DES VALEURS ABERRANTES	141
4.2.6 INTERPRETATION DES RESULTATS	143
4.2.7 OPTIMISATION DU CALCUL DES DISTRIBUTIONS NULLES	145
4.3 Application pratique	145
4.4 Conclusion	150
Discussion générale	151
Conclusion générale	154
Références	156
Annexes	182

Liste des figures

Figure 1 : Probabilité de survie des patients au cours du temps suivant leur année de naissance

Figure 2 : Représentation des atteintes les plus courantes dans la mucoviscidose

Figure 3 : Représentation schématique de la protéine CFTR

Figure 4 : Schéma simplifié du processus d'ouvrir du canal chlorure

Figure 5 : Les différentes étapes de maturation de la protéine CFTR

Figure 6 : Les 7 classes de mutation et leurs stratégies thérapeutiques associées

Figure 7 : Visualisation d'un spiromètre et représentation des VEMS (FEV1), CVF (FVC) et DEM 25-75

Figure 8 : Mécanisme de mesure du LCI

Figure 9 : Formule chimique du VX-770 (ivacaftor)

Figure 10 : Formule chimique du VX-809 (lumacaftor)

Figure 11 : Formule chimique du VX-661 (tezacaftor)

Figure 12 : Formule chimique du VX-445 (Elexacaftor)

Figure 13 : Fonctionnement du TIMS

Figure 14 : Visualisation de l'effet de l'ACP dans un cas à deux dimensions.

Figure 15 : Visualisation de l'ACP dans le cas d'une réduction de dimension de 3 à 2 composantes.

Figure 16 : Robustesse de l'ACP vis-à-vis de la taille de la cohorte

Figure 17 : Robustesse de l'ACP vis-à-vis de l'imputation des données manquantes

Figure 18 : Cercle de corrélation entre les paramètres cliniques et les trois premières composantes de l'ACP par les patients de 12-18 ans

Figure 19 : Projection des patients de 12-18 ans dans l'espace PC1-PC2

Figure 20 : Projection des patients de 12-18 ans dans l'espace PC1-PC2-PC3

Figure 21 : Distribution des paramètres dans les 4 groupes (clusters) pour les patients de 12-18 ans

Figure 22 : Cercle de corrélation entre les paramètres cliniques et les deux premières composantes de l'ACP pour les patients de 6-12 ans

Figure 23 : Composantes 1 et 2 de l'analyse par ACP des données protéomiques des 78 patients de l'étude protéomique.

Figure 24 : Graphique en volcan des protéines différentiellement exprimées entre les patients atteints de mucoviscidose et les individus sains

Figure 25 : Heatmap des protéines différenciellement exprimées entre les patients atteints par la mucoviscidose et les contrôles sains

Figure 26 : Nombre de protéines détectées pour les 10 patients traités par la trithérapie

Figure 27 : Nombre de protéines détectées lors d'une analyse différentielle à partir des échantillons de 10 patients avant et après le début du traitement par la trithérapie.

Figure 28 : Résumé du processus de culture cellulaire

Figure 29 : Liste des voies différenciellement exprimées chez les patients homozygotes pour la mutation F508del selon qu'ils soient ou non traités par VX-770/VX-661/VX-445

Figure 30 : Représentation des matrices d'intérêt pour la réalisation d'une ACP. Visualisation des données exploitées par rRoma.

Figure 31 : Activité des échantillons pour les deux voies déplacées

Figure 32 : Activité des échantillons pour la voie surdispersée

Figure 33 : Gènes aux poids les plus importants pour la voie surdispersée

Figure 34 : Comparaison des scores d'activité entre les deux groupes étudiés pour les collections surdispersées ou déplacées

Liste des annexes

Annexe 1 : Patients characteristics at baseline and at 6 months lumacaftor-ivacaftor using imputed ITT data. Values in n(%) or mean (SD). p values are calculated by paired t-test, or Wilcoxon signed rank test when comparing 6-12 to 12-18 year-old children.

Annexe 2 : Correlation coefficients and corresponding p values between the 6 outcome parameters and the 6 Principal Components for patients aged 12 years and older. Percentage of explained variance for each PC.

Annexe 3 : Loading matrix for the 6 outcome parameters and the 6 Principal Components for patients aged 12 years and older.

Annexe 4 : Mean (min/max) parameter values according to patient clustering in patients aged 12 years and older, and comparison of these clusters according to the studied outcome parameters.

Annexe 5 : Correlation coefficients and corresponding p values between the 6 outcome parameters and the 6 Principal Components for 6-12 year-old patients. Percentage of explained variance for each PC.

Annexe 6 : Loading matrix for the 6 outcome parameters and the 6 Principal Components for 6-12 year-old patients.

Annexe 7 : Patients' characteristics at the moment of sweat collection

Annexe 8 : List of differentially expressed proteins in CF and WT sweat.

Annexe 9 : Differentially expressed proteins in delF508 cells treated or not with the tritherapy

Annexe 10 : List of Overdispersed and Shifted Modules according to rRoma

Articles publiés et preprint

Liste des abréviations

ABC : Cassette liée à l'ATP

ACP : Analyse en Composantes Principales

ADNc : ADN complémentaire

ARNm : ARN messenger

ATB : nombre de jours d'antibiotiques dans les 6 derniers mois pour traiter les exacerbations pulmonaires

ATP : Adénosine TriPhosphate

CFLD : Cystic Fibrosis Liver Disease

CFRD : Cystic Fibrosis-related diabetes

CFTR : Cystic Fibrosis Transmembrane Regulator

CQRE : contrôle de qualité du RE

CRCM : centre de ressources et de compétences de la mucoviscidose

CVF : Capacité Vitale Forcée

DEM 25-75 : Débit Expiré Maximal entre les valeurs de 25 % et 75 % de la capacité vitale forcée

dNTP : désoxynucléoside triphosphate

domaine R : domaine Régulateur

EMA : European Medicines Agency

F508del : mutation p.Phe508delCFTR

FDA : Food and Drug Administration

FDR : Taux de Fausses Découvertes

GSEA : Gene Set Enrichment Analysis

HNE : cellules épithéliales nasales humaines

HPLC : chromatographie en phase liquide à haute performance

HTS : High throughput screening

IMC : Indice de Masse Corporelle

LC-MS : Chromatographie en phase liquide - spectrométrie de masse

LCI : Indice de clairance pulmonaire

MAR : Manquant de manière aléatoire

MCAR : Manquant de manière complètement aléatoire

MNAR : Manquant de manière non aléatoire

MSD1 : Domaine d'extension membranaire 1
MSD2 : Domaine d'extension membranaire 2
NBD1 : Domaine de liaison des nucléotides 1
NBD2 : Domaine de liaison des nucléotides 2
PC : Composante Principale
PCR : réaction en chaîne par polymérase
pp : pourcentage prédit
RE : réticulum endoplasmique
RNA-seq : Séquençage ARN
S-Trap : suspension trap
SOID : syndrome d'obstruction intestinale distale
ST : Test de la Sueur
TIMS : Trapped Ion Mobility Spectrometry
TIR : trypsine immuno-réactive
MSD1 : domaine transmembranaire 1
MSD2 : domaine transmembranaire 2
TOF : Time Of Flight
VEMS : Volume Expiré Maximal en une Seconde
VX-445 : Elexacaftor
VX-661 : Tezacaftor
VX-661/VX-770 : Symkevi
VX-770 : Ivacaftor
VX-770/VX-661/VX-445 : Trikafta
VX-770/VX-809 : Orkambi
VX-809 : Lumacaftor

I. Introduction

1. La mucoviscidose

1.1 Population touchée

La Mucoviscidose (en anglais Cystic Fibrosis) est une maladie héréditaire autosomique récessive provoquée par une mutation du gène Cystic Fibrosis Transmembrane Regulator (CFTR). Il s'agit d'une maladie rare (NORD's Rare Disease Database), c'est-à-dire d'une maladie touchant moins de 200000 personnes aux Etats-Unis, ou ayant une prévalence inférieure à 5 pour 10000 en Europe. On compte plus de 7000 maladies rares, touchant 30 millions de personnes aux Etats-Unis, et le même nombre en Europe. Parmi elles, la mucoviscidose est la plus fréquente des maladies récessives dans la population caucasienne, avec un cas toutes les 3000 naissances environ (Walters 2007). La prévalence varie cependant entre les pays (Mehta 2010), et fortement entre les ethnies (Walters 2007). Elle est par exemple de 1 pour 3000 parmi les Américains d'origine caucasienne, de 1 pour 4000 à 10000 parmi ceux d'origine latine, et de 1 pour 15000 à 20000 parmi ceux d'origine africaine. La maladie est encore plus rare en Asie, la fréquence de cas de mucoviscidose étant de seulement 1 pour 350000 au Japon (Yamashiro 1997).

En France, l'incidence varie selon les régions, de 1/2500 dans le nord-ouest à 1/10000 dans le sud-est. On recensait 7376 patients en 2020, dont près de 60 % d'adultes, alors que ces derniers ne constituaient que 18 % des patients atteints de la mucoviscidose en 1992 (Registre Français de la Mucoviscidose 2020). Ces chiffres témoignent de l'amélioration drastique des prises en charge qu'ont connu les patients au cours des deux dernières décennies, liée au développement du dépistage néo natal ainsi qu'à la mise au point de nouvelles thérapies. L'espérance de vie des patients a énormément augmenté au cours du dernier siècle ; de moins de 5 ans dans les années 50, elle est aujourd'hui de plus de 40 ans (Lopes-Pacheco 2016), et devrait encore augmenter grâce aux nouveaux traitements apparus dans les années 2010, comme le montre la Figure 1: la probabilité de survie des patients nés après 2000 est à tout âge supérieure à celle des patients nés dans les années 1990.

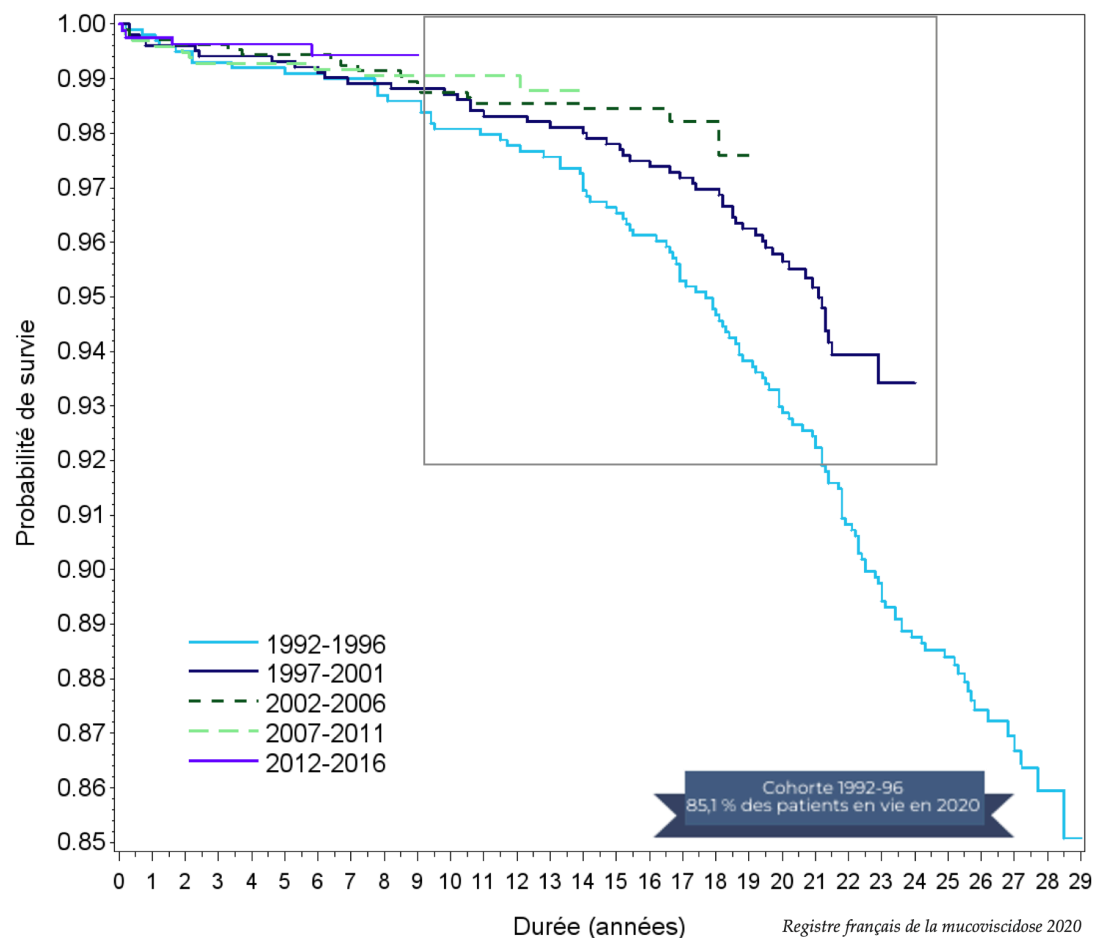


Figure 1: Probabilité de survie des patients (ordonnée) en fonction de leur âge (abscisse). Chaque courbe correspond à une période de date de naissance différente, entre 1992 et 2016.
(Registre Français de la Mucoviscidose 2020)

1.2 Diagnostic de la maladie

Depuis 2002, un dépistage néonatal systématique est réalisé en France pour diagnostiquer la mucoviscidose (Pollitt 1998). Le dépistage consiste à doser la trypsine immuno-réactive (TIR) dans des gouttes de sang séchées du nouveau né. La TIR est une protéine surexprimée en cas d'anomalie pancréatique, ce qui est le cas des personnes atteintes de mucoviscidose. La valeur mesurée est considérée positive si supérieure à 65yg/L. Cependant ce marqueur n'est pas totalement spécifique de cette maladie, et peut aussi sélectionner des enfants sains. Il est donc nécessaire de confirmer le diagnostic par une recherche ADN des mutations CFTR les plus fréquentes (parmi 29 mutations) (Registre Français de la Mucoviscidose 2020).

Trois cas de figure peuvent alors se présenter (Haute Autorité de Santé 2020) :

- deux mutations sont identifiées. Le nouveau-né et ses parents sont convoqués dans un centre de ressources et de compétences de la mucoviscidose (CRCM) pour une confirmation du diagnostic par la mise en place d'un test de la sueur.
- une seule mutation est identifiée. On réalise un test de la sueur comme précédemment. Si le test est positif, l'enfant est diagnostiqué atteint de la maladie. A l'inverse, s'il est négatif, l'enfant sera considéré hétérozygote et l'information transmise aux parents.
- aucune mutation n'est retrouvée, mais la TIR est très élevée. Dans ce cas, on mesure à nouveau la concentration de la TIR à 21 jours. Dans le cas où elle serait toujours élevée (le seuil est à 40 yg/L à 21 jours), un test de la sueur est réalisé.

La sueur des patients atteints par la mucoviscidose est très concentrée en ions chlorure (Di Sant'Agnese 1953). Le test de la sueur consiste à tirer avantage de ce fait pour diagnostiquer la maladie en dosant les ions chlorure présents dans la sueur des enfants après stimulation de la sudation (Sermet 2010). Un test est considéré positif dès lors que la valeur dépasse 60mmol/L (Shwachman 1981). Il est considéré « intermédiaire » si le résultat est compris entre 30 et 60mmol/L. Dans un tel cas, il est difficile de conclure à un diagnostic clair et le test de la sueur doit être réalisé à nouveau. S'il reste intermédiaire, des examens complémentaires peuvent être réalisés, notamment en mesurant la différence de potentiel nasal trans-épithéliale in vivo. Cette différence de potentiel vient d'un échange d'ions à travers l'épithélium et est donc modifiée dans le cas de la mucoviscidose (Farrell 2008).

Si l'étude génétique n'est pas réalisée (notamment en cas de refus des parents) alors que la TIR était anormalement élevée, le protocole est identique au cas où aucune mutation n'est retrouvée lors du test ADN.

D'autres signes cliniques permettent d'orienter le diagnostic de mucoviscidose, notamment : une faible prise de poids, des selles grasses et abondantes, des difficultés à digérer, des toux fréquentes ainsi que des fonctions respiratoires altérées (Davies 2007).

1.3 Symptômes

La mucoviscidose touche de multiples organes. Les atteintes les plus sévères sont situées au niveau du foie, des intestins, du pancréas et des voies respiratoires (poumon principalement). Ce sont les dérégulations dans ces dernières qui sont le plus souvent responsables de la mort des

patients, par insuffisance respiratoire. Mais l'on note aussi des défauts au niveau de la peau et des parties génitales (Figure 2).

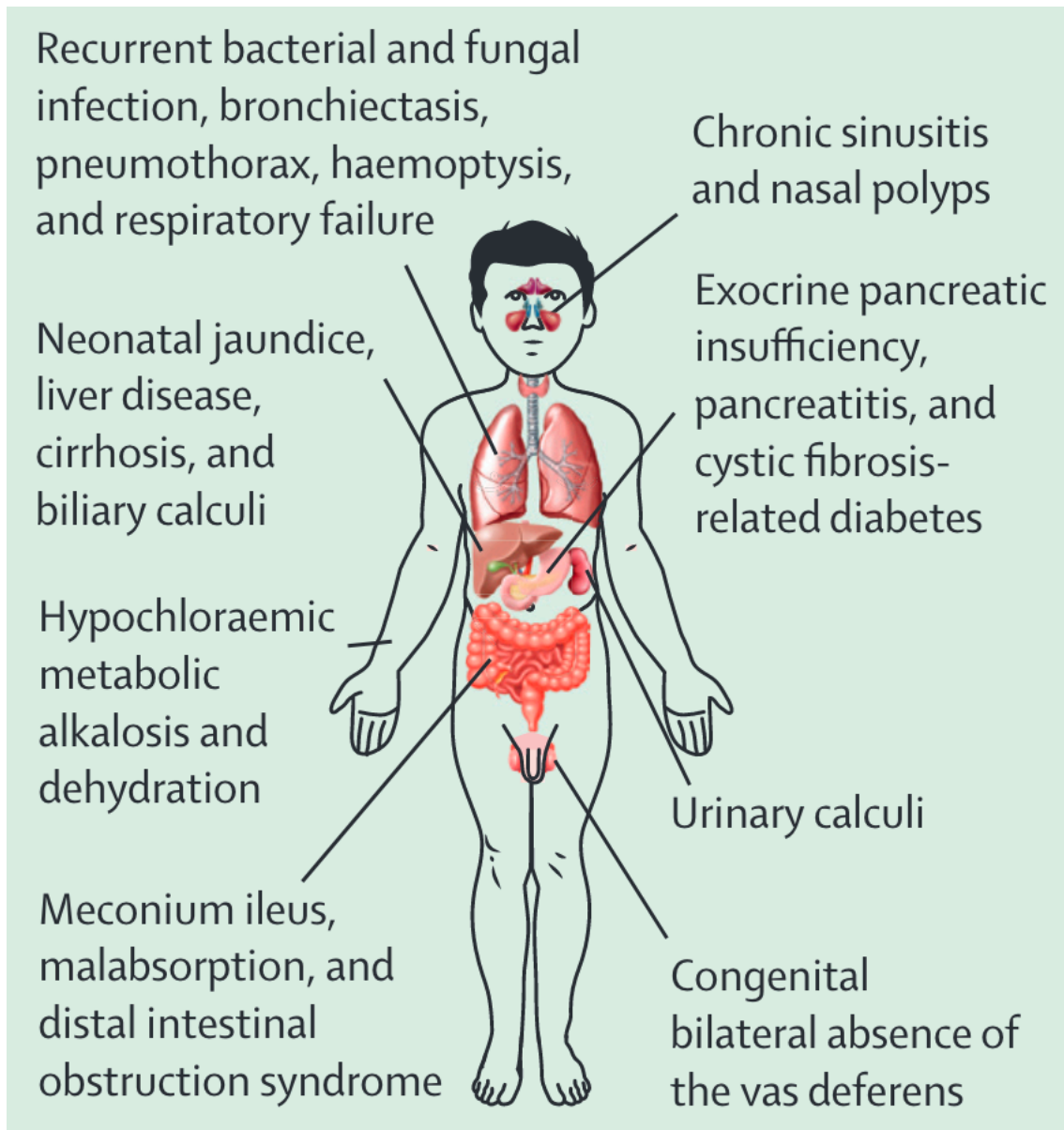


Figure 2 : Représentation des atteintes les plus courantes dans la mucoviscidose. Les atteintes principales sont retrouvées au niveau du nez, des poumons, du foie, du pancréas, de la peau, des intestins, et des parties génitales.

(Shteinberg 2021)

1.3.1 SYMPTÔMES PULMONAIRES ET RESPIRATOIRES

Les symptômes les plus fréquents dans la mucoviscidose sont des atteintes chroniques des voies respiratoires : obstructions bronchiques, inflammation et infections chroniques (Sly 2009).

Dans les poumons, les mutations du gène CFTR perturbent la sécrétion d'ions chlorure. En parallèle, une surabsorption de sodium est observée. Cela déshydrate le liquide de surface épithélial (Matsui 1998), pourtant important pour les fonctions ciliaires et antibactériennes. Le mucus s'épaissit, rendant plus difficile sa clairance, et favorisant le développement de bactéries en son sein (Boucher 2007). Les poumons sont progressivement endommagés par ces contraintes et leurs capacités fonctionnelles déclinent rapidement sans traitement, pour mener à des bronchectasies et une faillite respiratoire.

Si l'inflammation et l'infection des voies respiratoires sont deux facteurs connus de la mucoviscidose, il n'est toujours pas clairement établi lequel des deux précède l'autre. Si des données semblent montrer que l'inflammation peut exister sans infection (Esther 2019), l'on sait aussi que ces infections se développent chez les patients dès leur plus jeune âge (Lund-Palau 2016) notamment par *Staphylococcus aureus*.

1.3.2 SYMPTÔMES PANCRÉATIQUES

Environ 85% des enfants atteints de la mucoviscidose souffrent d'une insuffisance pancréatique. Ce pourcentage est plus élevé chez les patients porteurs de mutations associées à des formes plus sévères de la maladie (Singh 2017). Dans le pancréas, l'absence de CFTR est liée à une augmentation de la viscosité du mucus, causant une accumulation d'enzymes du pancréas, elle-même à l'origine d'une obstruction progressive des canaux pancréatiques, de la formation de kystes et de la dégradation de l'organe (Castellani 2017). C'est d'ailleurs de ces caractéristiques que vient le nom de la maladie en anglais : Cystic Fibrosis of the pancreas, pour fibrose kystique du pancréas. Les enfants ayant une insuffisance pancréatique souffrent de plus d'un mauvais état nutritionnel et d'une carence en vitamines A, D, E, K qui doivent être apportées par des traitements.

1.3.3 SYMPTÔMES INTESTINAUX

Les atteintes digestives sont présentes chez tous les patients, quelle que soit la mutation. 20% des patients souffrent d'un iléus méconial à la naissance. L'iléon terminal est obstrué par du méconium visqueux, pouvant entraîner dans les cas les plus graves une perforation de ce dernier, et une septicémie (Sathe 2017). Il s'agit le plus souvent de la première manifestation de la mucoviscidose chez les patients.

Chez les adultes, on note une incidence accrue des cas de reflux gastro-oesophagiens (Demeyer 2016). 15% des patients souffrent aussi d'un syndrome d'obstruction intestinale distale

(SOID), conséquence d'une obstruction de l'intestin par une accumulation de matériel visqueux. Elle est similaire à l'iléus méconial du nouveau-né. Les patients ayant souffert d'un iléus méconial sont d'ailleurs plus susceptibles d'avoir un SOID au cours de leur vie (Shteinberg 2021).

1.3.4 SYMPTÔMES DU FOIE

La Cystic Fibrosis Liver Disease (CFLD) est la troisième cause de mortalité chez les patients atteints par la mucoviscidose (Toledano 2019). Elle est détectée dans environ 20% des patients au cours de leur vie. Dans un fonctionnement normal, CFTR régule le flux de bile dans le foie. Lorsque la protéine est dysfonctionnelle, la bile visqueuse cause des obstructions dans cet organe, pouvant mener progressivement à de l'hypertension portale et à une cirrhose (Castellani 2017). L'évolution vers cette dernière est reportée dans 5 à 10% des patients au cours de leur première décennie de vie (Debray 2011).

1.3.5 SYMPTÔMES GÉNITAUX

Environ 95% des hommes atteints par la mucoviscidose sont infertiles du fait d'une agénésie bilatérale congénitale des canaux déférents (de Souza 2018). Ils ne sont cependant pas stériles, car la production de sperme reste normale dans les testicules pour la majorité des patients, qui peuvent donc avoir des enfants par reproduction assistée (McCallum 2000).

Les femmes peuvent présenter une hypofertilité du fait de l'hyperviscosité de la glaire cervicale (Kotloff et al. 1992), mais une grossesse reste tout de même possible.

1.3.6 SYMPTÔMES DE LA PEAU

Des cas de kératodermie palmoplantaire aquagénique (KPA) ont été observés et toucheraient environ 40% des patients (Tolland 2010, Garçon-Michel 2010). Ils se caractérisent par la formation rapide d'œdèmes ridés, épais et blanchâtres au niveau des paumes des mains et des plantes des pieds après immersion dans de l'eau. Ils s'accompagnent de prurit et peuvent être douloureux pour le patient. La cause de ces symptômes dans la mucoviscidose n'est pas encore comprise.

1.3.7 AUTRES SYMPTÔMES

Environ 35% des adultes souffrent de diabète appelés Cystic Fibrosis-related diabetes (CFRD), qui peuvent être de type 1 ou 2, et présenter différents niveaux de résistance à l'insuline (Olesen 2020).

Les patients ont aussi plus de difficultés à prendre du poids que des personnes saines, et leur masse osseuse est inférieure à indice de masse corporelle égal. Des manifestations osseuses (Cystic Fibrosis Related Bone Disease) apparaissent chez 20 à 35% des patients du fait d'une mauvaise absorption de la vitamine D (Putman 2019).

2. CFTR

2.1 Le gène CFTR

Le gène responsable de la mucoviscidose a été localisé en 1985 dans une région d'un million de paires de bases sur le chromosome 7 (Tsui 1985). Si l'hypothèse d'une maladie monogénique apparaissait de plus en plus plausible à cette époque, elle n'était pas totalement validée.

Quatre ans plus tard, en 1989, le gène CFTR fut identifié par une stratégie de clonage positionnel (Rommens 1989, Riordan 1989, Kerem 1989). Une carte physique de la région pressentie pour contenir le gène responsable de la mucoviscidose a dans un premier temps été réalisée par assemblage de clones, de courtes chaînes d'ADN. Puis les zones de cette carte contenant les gènes ont été identifiées, avant finalement de comparer les séquences des gènes entre individus sains et patients pour déterminer sur le quel d'entre eux spécifiquement se situe la mutation. Il a ainsi été déterminé que CFTR est un gène situé dans la région 7q31.2, long d'environ 190kb d'ADN génomique humain, transcrit en un ARNm de 6.2kb (Collins 1992). Il code pour la protéine du même nom, longue de 1480 acides aminés.

Le gène CFTR est principalement exprimé dans le pancréas, les intestins, les glandes sudoripares (qui sécrètent la sueur) et l'appareil respiratoire (The Human Protein Atlas). Dans les poumons, il a longtemps été supposé que les cellules ciliées étaient responsables de la majorité de la production de CFTR. Cependant, le développement du séquençage ARN (en anglais RNA sequencing, RNA-seq) de cellules uniques ont permis de montrer que ces cellules ne produisaient en réalité que de très petites quantités de cette protéine. La production est en majorité assurée par les cellules sécrétrices et basales. La protéine est aussi produite en grande quantité par les ionocytes, mais ces derniers sont rares dans les épithéliums respiratoires (Okuda 2020).

2.2 La protéine CFTR

La protéine CFTR, de poids moléculaire de 168 KiloDaltons dans sa forme mature (Estabrooks 2020), est constituée de 1480 acides aminés. Elle appartient à la famille des protéines ABC (ATP-Binding Cassette), une grande famille de protéines réalisant le transport de substrats à l'intérieur et à l'extérieur des cellules en hydrolysant l'Adénosine TriPhosphate (ATP). L'autre nom

de la protéine CFTR est d'ailleurs ABCC7, pour ATP-Binding Cassette sub-family C member 7. Sa forme fonctionnelle est localisée au niveau de la membrane apicale des cellules épithéliales, où elle fonctionne comme un canal ions chlorure.

Comme la plupart des protéines ABC, CFTR se compose de deux domaines transmembranaires (MSD1 et MSD2, pour membrane spanning domain 1 et 2) et de deux domaines de liaison à l'ATP dans le cytoplasme (NBD1 et NBD2, pour Nucleotide Binding Domain 1 et 2). Elle se distingue cependant des autres protéines ABC par la présence d'un domaine de régulation cytoplasmique (R, pour Regulatory), situé des résidus 580 à 831, qui sépare la partie MSD1-NBD1 de la partie MSD2-NBD2. De plus, contrairement à ses homologues qui sont pour la plupart des transporteurs actifs, CFTR agit de manière passive en formant un canal ionique transmembranaire (Csanady 2018). La figure 3 résume la structure de CFTR.

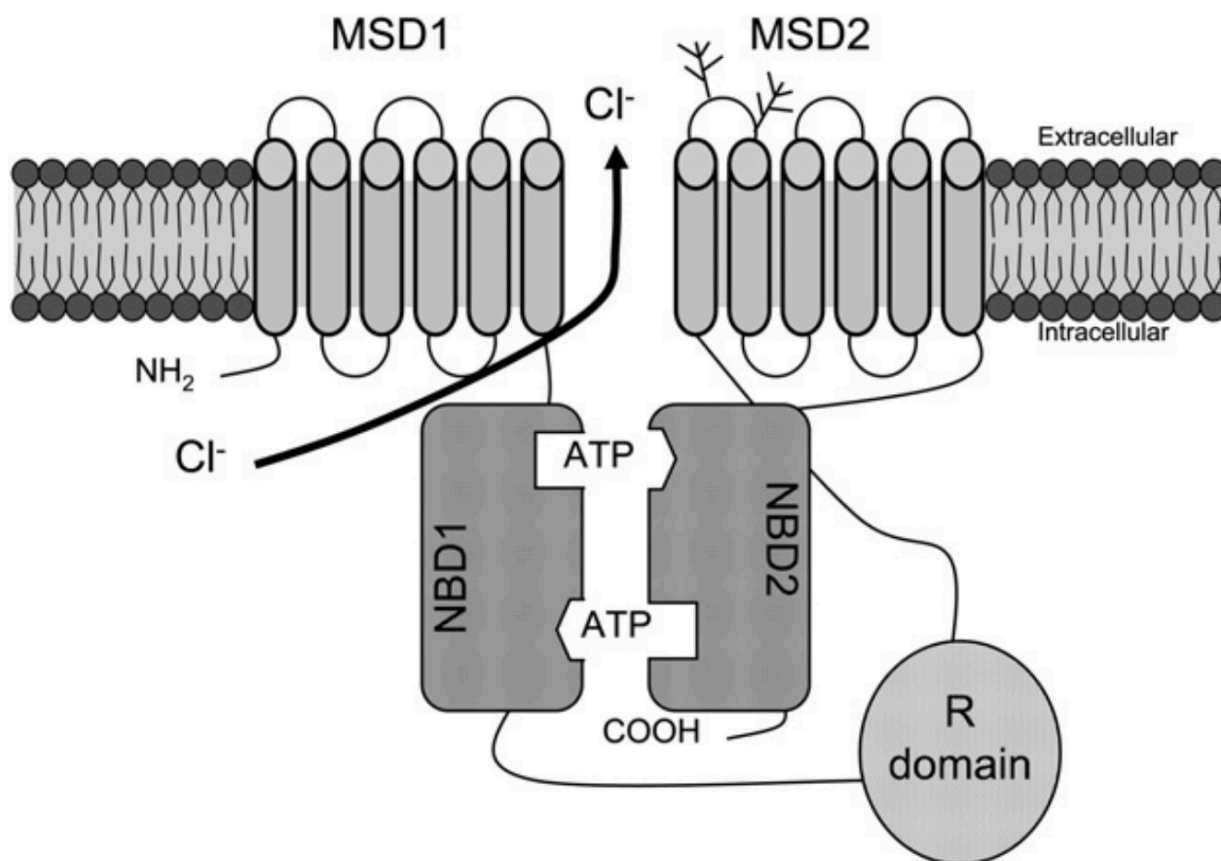


Figure 3 : Représentation schématique de la protéine CFTR (Rogan 2011). CFTR possède 5 domaines : deux domaines transmembranaires (MDS1 et MSD2), qui traversent la membrane ; deux domaines de liaison à l'ATP (NBD1 et NBD2), et un domaine régulateur (R). CFTR agit comme canal à ions chlorure ouvert lorsque des molécules d'ATP se lient aux sites NBD1 et NBD2.

Chacun des domaines transmembranaires est formé de 6 hélices hydrophobes permettant la formation du canal ions chlorure.

L'activation du canal ions chlorure de CFTR est dépendante de plusieurs éléments. Dans un premier temps, le domaine cytoplasmique R doit être phosphorylé par la protéine kinase A ou C (PKA et PKC) (Picciotto 1992). Tant que le domaine R n'est pas phosphorylé, le canal reste fermé, les étapes suivantes nécessaires à son ouverture ne pouvant pas avoir lieu (Bozoky 2013).

Dans un second temps, des molécules d'ATP viennent se lier aux sites NBD1 et NBD2 (Anderson 1991). Les deux régions forment un dimère, dont les zones d'interface sont les sites de liaison de l'ATP (Lewis 2004). Les domaines NBD interagissant avec les quatre boucles intracellulaires des MSD, la formation du dimère entraîne un changement de la conformation de ces derniers et l'ouverture du canal. L'énergie de l'hydrolyse de l'ATP permet la régulation de l'ouverture et de la fermeture du canal (Akabas 2000). En particulier, l'hydrolyse de l'ATP sur le site de liaison de NBD2 entraîne une fermeture rapide de ce dernier (Gadsby 2006).

La figure 4 résume le processus d'ouverture du canal ions chlorure.

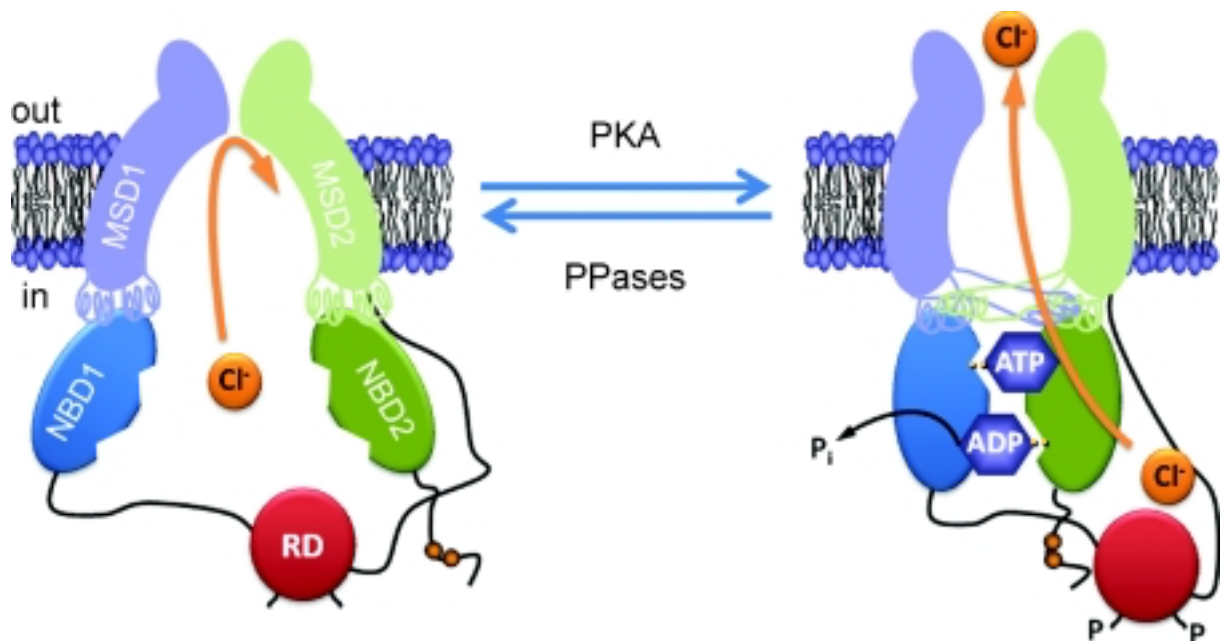


Figure 4 : Schéma simplifié du processus d'ouvrir du canal ions chlorure (Hwang 2009). Deux conditions sont nécessaires à l'ouverture du canal : le domaine régulateur (RD) doit être phosphorylé, et deux molécules d'ATP doivent se lier aux sites NBD1 et NBD2 pour que ces derniers forment un dimère menant à un changement de conformation des MSD1 et MSD2.

2.3 Biogénèse de CFTR

La synthèse de la protéine CFTR débute dans le noyau, par la transcription de l'ADN en ARN messager (ARNm). Ce dernier rejoint ensuite les ribosomes cytosoliques pour débiter la phase de traduction. La biogénèse suit alors plusieurs étapes au niveau du RE (Pranke 2014).

Le domaine transmembranaire MSD1 est tout d'abord positionné dans la membrane du RE. Le domaine NBD1 est ensuite synthétisé, et stabilisé par des protéines chaperones Hsp40 et Hsp70 en parallèle de la synthèse du domaine R. Ce dernier, une fois synthétisé, interagit avec le domaine N-terminal de la protéine, menant à la séparation d'une partie des protéines chaperones de CFTR. C'est ensuite le site MSD2 qui est synthétisé, amené à la membrane du réticulum endoplasmique (RE), et associé à MSD1 pour former le complexe MSD1-MSD2, ce qui a aussi pour effet de stabiliser le domaine NBD1-R et de séparer la plupart des protéines chaperones de ce complexe. Enfin, le domaine NBD2 est synthétisé, et des repliements post-traductionnels sont effectués, impliquant des protéines chaperones HSP40, HSP70 et HSP90 (Farinha 2002), et la calnexine (Okiyoneda 2008). Les peptides migrent ensuite vers l'appareil de Golgi pour être glycosylés et ainsi former la protéine mature.

Durant la totalité de leur temps de résidence dans le RE, les peptides de CFTR sont sujets au système de contrôle de qualité du RE (CQRE). Le CQRE s'assure que seules les protéines correctement repliées atteignent leur maturation et puissent migrer vers leur destination finale dans la cellule. A l'inverse, les protéines mal repliées sont dirigées vers les voies de dégradation. Le rôle du CQRE et des protéines associées n'est cependant pas totalement élucidé. Des données *in vivo* suggèrent une amélioration du repliement de CFTR grâce aux protéines chaperones (Meacham 1999) alors que des données dans la levure suggèrent l'inverse (Youker 2004).

Le défaut de repliement entraîne d'autres perturbations cellulaires, notamment un stress du RE qui mène à l'activation de l'Unfolded Protein Response, un processus qui ralentit la traduction des protéines et favorise leur dégradation (Bartoszewski 2008).

Même dans le cas d'absence de mutation dans le gène CFTR, une part non négligeable des protéines n'atteignent pas l'appareil de Golgi, car elles sont préalablement ciblées par le mécanisme de dégradation du RE (Ward 1994). Cette proportion semble cependant varier entre 0 et 80% en fonction du type cellulaire (Varga 2004). A l'inverse, on peut retrouver une petite quantité de protéines CFTR au niveau de la membrane pour certaines mutations de ce gène. C'est notamment le cas pour la plus courante d'entre elles, p.Phe508delCFTR (F508del) (Sermet-Gaudelus 2002). Ces observations montrent qu'il est possible de transporter une protéine mutée à la membrane, et

justifie la recherche de traitements visant à l'amélioration de la maturation et du transport de cette dernière.

La figure 5 résume le processus de maturation de la protéine CFTR.

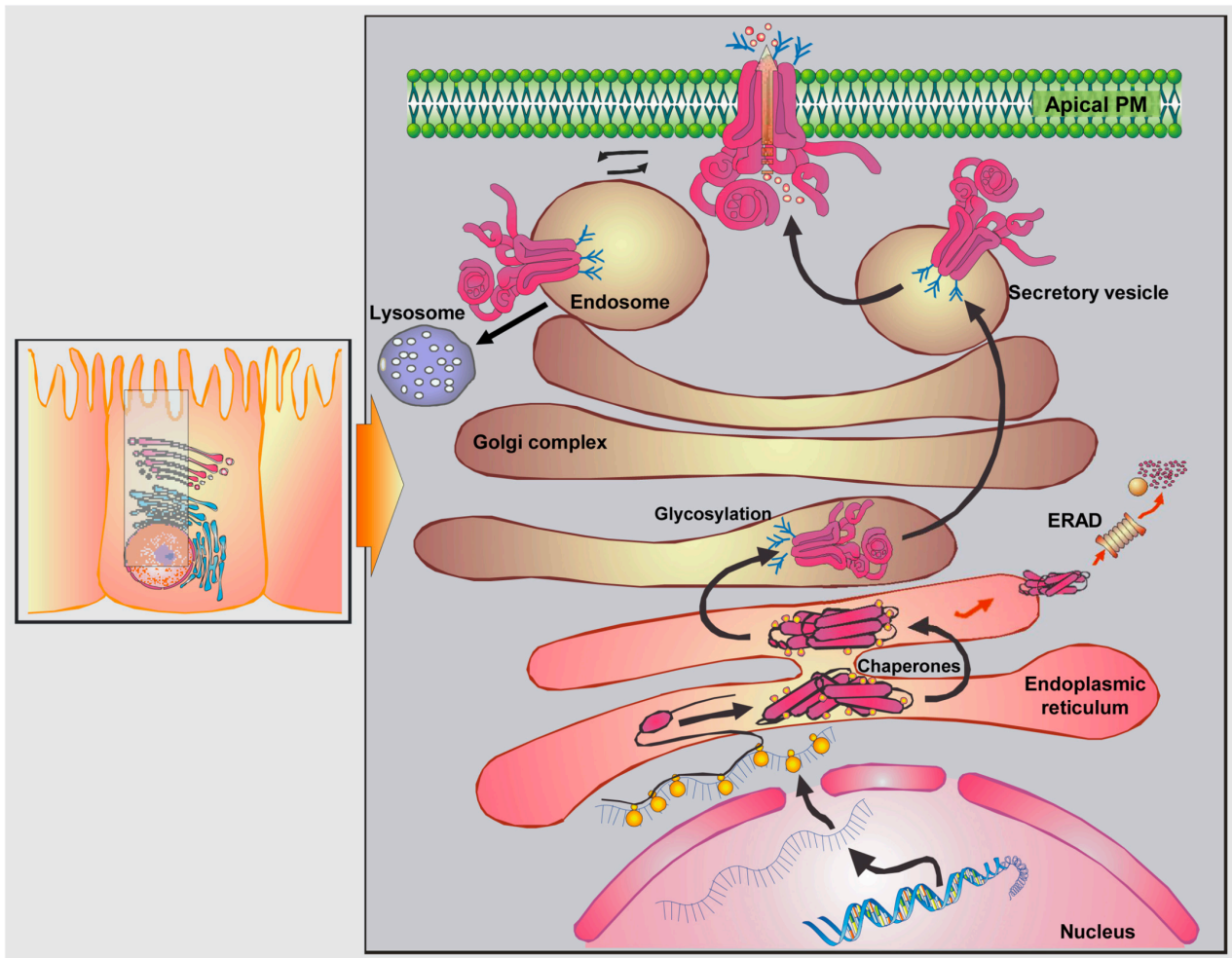


Figure 5: Les différentes étapes de maturation de la protéine CFTR (Pranke 2014). Les différentes parties de la protéine sont d'abord synthétisées séparément, puis stabilisées par des chaperones dans le RE. Les peptides ainsi produits sont soumis au mécanisme de contrôle de qualité du RE et dégradés si considérés non conformes. Dans le cas contraire, ils migrent dans l'appareil de Golgi afin d'être glycosylés pour former la protéine mature. Cette dernière migre ensuite jusqu'à la membrane plasmique via des vésicules sécrétoires.

2.4 Les classes de mutation

Une première classification des mutations du gène CFTR fut proposée dès 1993 par Welsh (Welsh 1993), qui souhaitait les regrouper en fonction de leurs défauts biologiques. Une classification en 6 groupes a plus tard été largement adoptée, séparant les mutations en fonction du type de défaut moléculaire et de la sévérité des défauts cellulaires causés. Ces dernières années, la considération d'une septième classe de mutation a été progressivement acceptée (Marson 2016). Les sept classes de mutation sont décrites dans les sous sections suivantes et résumées dans la figure 6 à la fin de cette section.

2.4.1 CLASSE I : MUTATION MENANT À L'ABSENCE DE SYNTHÈSE DE LA PROTÉINE MALGRÉ LA PRODUCTION D'ARNM

Les mutations de classe 1 mènent à la quasi absence de production de protéine CFTR. Elles incluent les mutations non-sens (remplacement d'un codon codant pour un acide aminé par un codon stop) et les mutations entraînant un décalage du cadre de lecture menant à l'apparition d'un codon stop prématuré. Les ARNm transcrits à partir de ces gènes mutés sont en grande majorité dégradés avant la traduction, entraînant une absence quasi totale de protéine dans la cellule.

Cette classe inclut la mutation G542X, la deuxième plus fréquente dans le monde, affectant 5.5 % des patients atteints par la mucoviscidose (Alonso 2007).

Dans la classification précédente, cette classe incluait aussi les mutations menant à l'absence de production d'ARNm pour CFTR (aujourd'hui mutations de classe VII).

2.4.2 CLASSE II : MUTATION AFFECTANT LA MATURATION DE LA PROTÉINE CFTR

Cette classe inclut la mutation F508del, la plus fréquente des mutations de CFTR, touchant 83% des patients français (registre Français de la Mucoviscidose 2020)

Chez les patients porteurs de mutations de classe II, l'ARNm est entièrement produit et la protéine synthétisée, cependant son repliement est défectueux. Ces défauts sont reconnus par des protéines chaperones dans le RE, entraînant la rétention de la protéine dans ce dernier pour être dégradée. CFTR n'atteint donc pas la membrane plasmique, et ne peut pas jouer son rôle de canal ionique.

Cette classe étant la plus fréquemment observée chez les patients, et entraînant des formes graves de la maladie, c'est sur cette dernière que sont portés la plupart des efforts de recherche de nouveaux traitements.

2.4.3 CLASSE III : MUTATION ENTRAÎNANT UN DÉFAUT DE LA FONCTION CANAL CHLORURE DE CFTR

Cette classe contient les mutations dites de « gating ». La protéine est produite et atteint la membrane cellulaire. Elle n'est cependant pas en mesure d'ouvrir le canal ions chlorure. La mutation la plus courante de cette classe est G551D, présente chez 2% des patients en France.

Ces mutations sont associées à des formes graves de la maladie (Bompadre 2007). Cependant, les traitements visant les mutations de classe II sont aussi efficaces sur celles de classe III.

2.4.4 CLASSE IV : MUTATION LIMITANT LA CONDUCTANCE DU CANAL CHLORURE DE CFTR

Pour cette classe de mutations la protéine atteint la membrane et parvient à jouer son rôle de canal ions chlorure. Néanmoins la conductance de ce dernier est réduite, limitant le flux d'ions.

Ces mutations, dont la plus commune est R117H, présente chez 2% des patients en France, sont associées à des formes moins sévères de la maladie, bien qu'entraînant l'infertilité des patients homme (Yu 2016, de Nooijer 2011).

2.4.5 CLASSE V : MUTATION LIMITANT LA PRODUCTION DE LA PROTÉINE CFTR

Dans cette classe l'ARNm produit est instable, réduisant fortement la quantité de protéine CFTR produite. Il s'agit principalement de mutations d'épissage ou induisant un saut d'exon.

Si la quantité de protéine CFTR produite est réduite, elle joue néanmoins correctement son rôle de canal ions chlorure, et le flux d'ions est normal.

La mutation de classe V la plus courante est A455G, présente chez 0.6% des patients en France. Elle est associée à une forme peu sévère de la maladie.

2.4.6 CLASSE VI : MUTATION MENANT À UNE PROTÉINE CFTR INSTABLE À LA MEMBRANE

Pour cette mutation, la quantité de protéine produite est normale, et elle agit correctement en tant que canal ions chlorure. Elle est cependant instable à la membrane, est recyclée prématurément à partir de cette dernière, et est dégradée dans les lysosomes (De Boeck 2020).

Il s'agit de la classe de mutations associée aux formes les moins sévères de la maladie.

2.4.7 CLASSE VII : MUTATION MENANT À L'ABSENCE DE PRODUCTION D'ARNM

Les mutations de cette classe correspondent à des grands réarrangements de l'ADN entraînant l'absence de production d'ARNm pour CFTR.

Cette classe de mutation était auparavant incluse dans les mutations de classe I (aucune protéine n'est produite), mais a été rajoutée pendant le développement des modulateurs de CFTR pour mieux caractériser les effets de ces derniers.

Les mutations de cette classe correspondent aux formes de la maladie les plus difficiles à traiter. Ni la protéine, ni l'ARNm, ne peuvent être sauvés par des thérapies correctrices (Marson 2016). La mutation 1717-1G>A est présente chez 2% des patients en France.

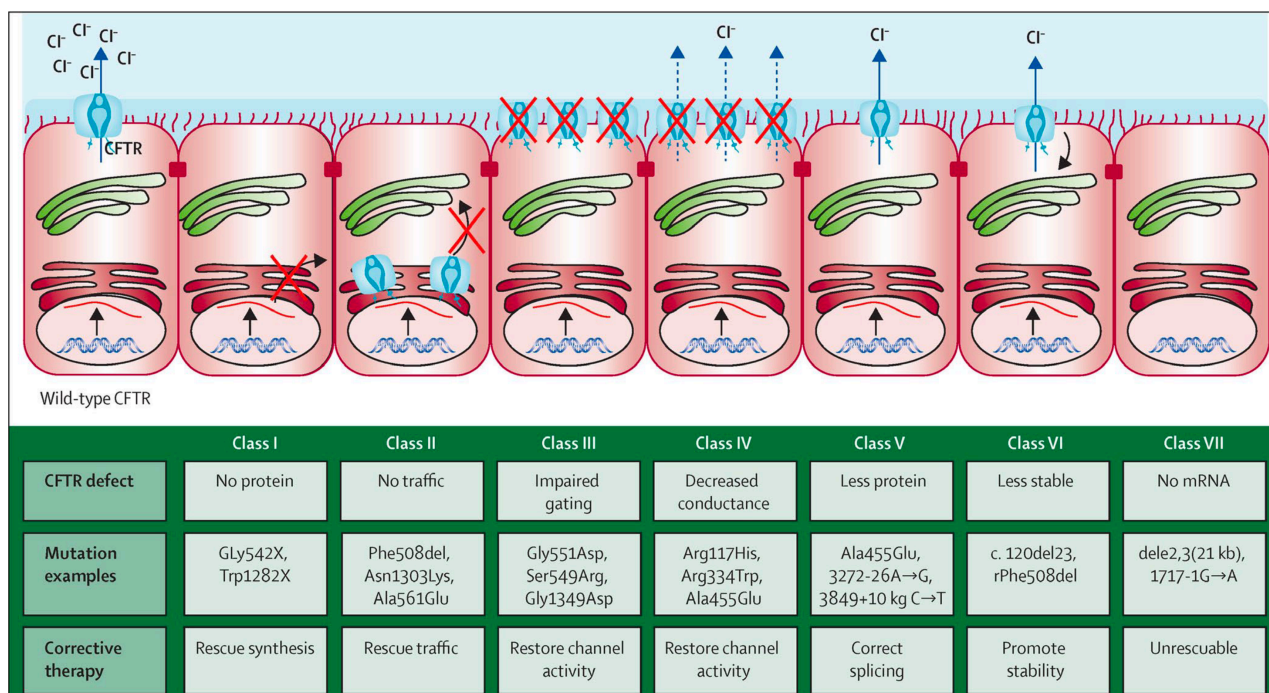


Figure 6 : Description des 7 classes de mutation et des stratégies thérapeutiques associées. Pour chaque classe de mutation, un schéma d'une cellule épithéliale met en évidence le défaut associé, aussi décrit dans le tableau situé en dessous, ainsi que les mutations les plus courantes.

(De Boeck 2020)

3. Traitement par modulateurs

La mucoviscidose est causée par des mutations sur un seul gène : CFTR. Comme on a pu le voir dans la section précédente, l'impact sur la protéine varie suivant la classe de la mutation. La production peut être inexistante ou réduite, et la fonction des protéines qui atteignent la membrane plasmique peut être altérée. Historiquement, les traitements pour la mucoviscidose ont tout d'abord été symptomatiques : lutte contre les infections chroniques par antibiotiques, contre les insuffisances pancréatiques par apport d'enzymes pancréatiques, contre la difficulté à prendre du poids par un suivi diététique (Davies 2007). Ces dernières années, la recherche dans la mucoviscidose a changé d'orientation pour se concentrer sur le développement de molécules, appelées modulateurs, qui permettraient de restaurer la production de protéine CFTR et son transport à la membrane, ainsi que d'améliorer son activité.

Le développement de ces traitements a été rendu possible par la meilleure compréhension des différents défauts induits par les mutations du gène CFTR sur la protéine. Les molécules développées sont ainsi réparties en deux fonctions principales, correspondant à ces défauts : les correcteurs et les potentiateurs. Les correcteurs limitent la séquestration de la protéine dans le RE et améliorent son transport à la membrane, tandis que les potentiateurs augmentent l'activité du canal ions chlorure. Depuis 2020, des amplificateurs, qui augmentent la production d'ARNm, et donc de protéine CFTR, commencent aussi à être testés en essais cliniques (Dukovski 2020).

Quel que soit le médicament développé, il doit ultimement être testé en conditions réelles sur des patients, et son efficacité doit être mesurée. Pour cela, il est nécessaire de déterminer un ou plusieurs paramètres dont l'évolution témoignerait d'une amélioration de la vie des patients.

Dans ce chapitre nous décrirons dans un premier temps les paramètres cliniques qui sont ou ont été utilisés comme référence pour évaluer l'efficacité d'un traitement, puis nous nous plongerons dans l'historique du développement des traitements par modulateurs, et terminerons en évoquant les nouveaux challenges autour de la mesure clinique des effets des modulateurs aujourd'hui.

3.1 Mesures en clinique pour évaluer l'effet des traitements

L'effet d'un traitement est mesuré par l'évolution d'un paramètre clinique d'intérêt, défini comme étant une variable dont l'évolution permet de détecter un effet bénéfique pour les patients

(Biomarkers Definitions Working Group 2001), en particulier en lien avec sa survie. Cependant, dans l'optique de réaliser des essais cliniques de durée raisonnable (l'European Medicines Agency -EMA- recommande une durée de 6 mois), l'amélioration de l'espérance de vie dans la mucoviscidose a rendu obsolète le paramètre de durée de survie des patients, pourtant mesure de référence pour évaluer l'efficacité d'un traitement.

Une autre mesure doit donc être utilisée. L'EMA (EMA 2009), comme la Food and Drug Administration (FDA) (FDA 2009), recommandent le recours à l'évaluation du Volume Expiré Maximal en une Seconde (VEMS) ou d'autres paramètres respiratoires, incluant notamment la Capacité Vitale Forcée (CVF) et le Débit Expiré Maximal entre les valeurs de 25 % et 75 % de la capacité vitale forcée (DEM 25-75). Ces paramètres respiratoires sont aujourd'hui les mesures de référence utilisées dans les études cliniques de phase III, dernière étape avant la mise sur le marché d'un traitement, au cours desquelles on vérifie l'effet bénéfique de ce dernier pour les patients. L'utilisation de ces paramètres respiratoires est basée sur le lien fort et établi entre la valeur du VEMS et la survie des patients (Corey 1997). Le VEMS est par ailleurs corrélé à la qualité de vie ressentie des patients (Dill 2013).

Les VEMS, CVF et DEM 25-75 sont trois mesures des fonctions respiratoires réalisées lors d'un test dit de spirométrie (Liou 2009). Ces trois mesures sont représentées sur la figure 7. Le VEMS correspond à la quantité maximale d'air que peut exhaler un patient en une seconde, après une inspiration maximale. Le CVF correspond à la quantité maximale d'air que peut exhaler un patient, sans limite de temps, et sans reprendre sa respiration. Enfin, le DEM 25-75 correspond au débit d'expulsion lorsque le patient a expulsé entre 25 et 75% de l'air qu'il est capable d'expulser au total. Ces trois mesures sont obtenues en faisant souffler le patient dans un spiromètre, dont sont équipés tous les centres dédiés à la mucoviscidose, et dont le protocole d'utilisation est standardisé (Miller 2005), permettant une grande fiabilité dans les données obtenues. En revanche, les patients doivent être suffisamment âgés pour pouvoir comprendre les instructions liées à l'appareil et l'utiliser correctement. Bien que des études aient montré la possibilité de réaliser des tests spirométriques robustes chez des enfants de moins de 5 ans (Marostica 2002), il apparaît que seulement une partie des patients de ces âges en sont véritablement capables (Crenesse 2001). Les valeurs de VEMS, CVF et DEM 25-75 sont ainsi difficiles à obtenir de manière fiable pour des patients de moins de 5 ans.

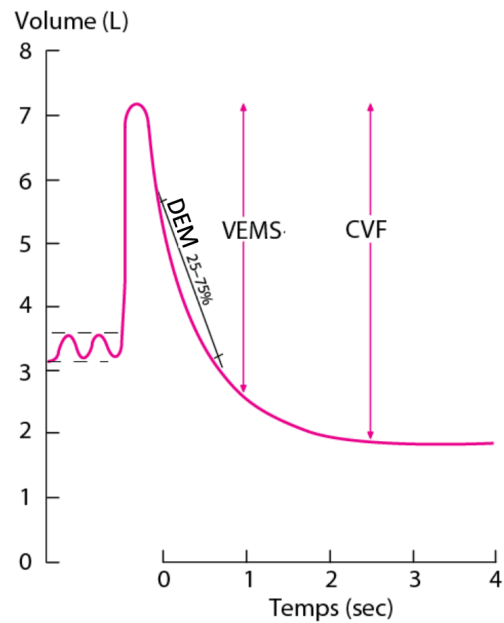


Figure 7 : Visualisation d'un spiromètre et représentation des VEMS, CVF et DEM 25-75. La courbe représentée correspond au volume d'air dans les poumons (en ordonnée). Le temps est donné en abscisse. Initialement, au repos, le volume d'air dans les poumons évolue peu. Le patient réalise ensuite une inspiration maximale, puis une expiration maximale. Le volume d'air expulsé au cours de la première seconde est le VEMS, et le volume d'air total expulsé est le CVF. Le DEM 25-75 est évalué en prenant le flux moyen entre 75% du volume d'air maximum dans les poumons et 25% du volume d'air maximum dans les poumons.

D'autres méthodes de mesure des fonctions pulmonaires ont donc été développées pour récupérer cette information chez les patients plus jeunes. La Raised Volume Thoracic Compression Technique existe depuis la fin des années 90 (The Joint American Thoracic Society/European Respiratory Society Working Group on Infant Lung Function 2000). Elle consiste à gonfler les poumons à une pression donnée pour imiter la capacité pulmonaire totale, puis à appliquer une pression sur le torse à l'aide d'un gilet gonflable pour forcer l'exhalation. Plutôt que le VEMS, qui est réalisé sur une seconde, c'est le volume d'air maximal expiré sur une demie seconde qui est mesuré, les patients plus jeunes ayant tendance à expulser plus vite l'air des poumons. Malheureusement, il est nécessaire d'avoir une cohorte de taille très grande pour mesurer des différences significatives entre patients atteints de la mucoviscidose et contrôles sains, bien que cela reste possible (Davis 2010). Il est donc très improbable de mesurer un effet significatif d'un traitement par cette méthode.

La méthode la plus prometteuse et la plus utilisée aujourd'hui comme alternative aux tests spirométriques chez les jeunes patients est le Lung Clearance Index (LCI). Il consiste à faire respirer au patient du dioxygène et à mesurer sa capacité à expulser le diazote contenu dans ses

poumons. Plus précisément, on mesure le nombre de respirations nécessaires pour que la concentration en diazote dans l'air exhalé ne soit plus que de un quarantième de la concentration initiale (80% dans l'air ambiant) (Stanojevic 2021). Le gaz de dioxygène pur se mélange facilement au gaz des poumons chez des personnes saines, et l'expulsion du diazote est donc réalisée en peu de respirations. A l'inverse, de petites obstructions des poumons, qui apparaissent dès le plus jeune âge chez les patients atteints de la mucoviscidose, peuvent piéger une partie du gaz et limiter la capacité de ce dernier à se mélanger. L'expulsion du diazote requiert donc plus d'expirations chez les patients atteints de la mucoviscidose. Le mécanisme de mesure est représenté sur la figure 8.

Cette mesure est très simple à réaliser chez de jeunes patients, car il suffit de placer le dispositif au niveau de leur bouche et de les distraire pour qu'ils continuent de respirer normalement (Robinson 2009). Elle possède de plus plusieurs autres avantages sur le VEMS. En effet, le LCI permet de distinguer les patients souffrant de la mucoviscidose de contrôles sains (Gustafsson 2003, Lum 2007), et ce de manière plus significative que le VEMS chez les patients de moins de 5 ans (Aurora 2005). De plus, cette mesure est corrélée à des dommages visibles sur des scanners à haute résolution, en particulier chez les plus jeunes enfants (Ramsey 2016).

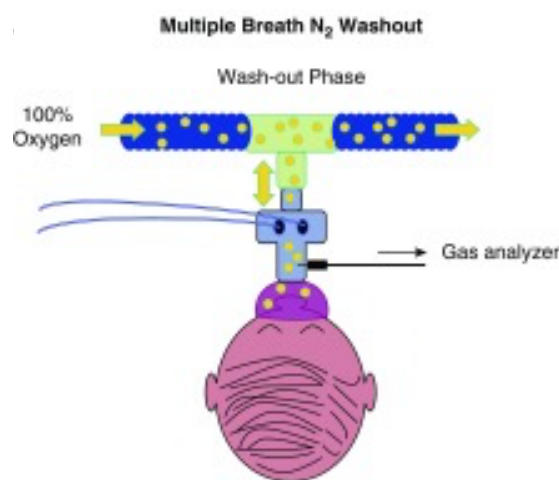


Figure 8 : Mécanisme de mesure du LCI (Subbarao 2015). Un dispositif est placé sur la bouche de l'enfant, et relié à une conduite dans laquelle circule du dioxygène pur. Un analyseur permet de mesurer la composition de l'air expiré par l'enfant.

Les scanners à haute résolution font par ailleurs partie des autres paramètres considérés. Plus précisément, ils sont utilisés afin de déceler des traces de défauts pulmonaires. Des radios sont obtenues pendant une grande inspiration, puis pendant une expiration totale (Judge 2006). A chaque image obtenue est ensuite associé un score prenant compte des différentes anomalies qui peuvent

exister dans les poumons et de leur niveau de sévérité (bronchectasie, mucus, opacités, air piégé) dans le but de normaliser les observations faites sur chaque image. Le score utilisé était initialement celui de Bhalla (Bhalla 1991), mais d'autres méthodes ont été proposées, dont le CF-CT score, le plus couramment utilisé aujourd'hui (Stick 2011). Cependant il a été montré que toutes ces méthodes sont robustes, reproductibles, et corrélées aux fonctions pulmonaires (De Jong 2004). Plusieurs études ont mis en lumière l'intérêt de ces scanners pour la prédiction des conditions futures des patients. Il a notamment été montré que le score dans l'enfance était prédicteur des valeurs de spirométrie à l'âge adulte (Sanders 2011), et qu'un score élevé prédisait des exacerbations pulmonaires (Sanders 2015) ainsi que des infections (Bortoluzzi 2014), et ce de manière plus précise que la mesure du VEMS. Cependant, la réalisation d'un scan est onéreuse.

Les traitements par modulateurs visant à corriger les défauts de CFTR, il est aussi possible de mesurer leurs effets en évaluant directement le rétablissement de la fonction canal ions chlorure de CFTR. Les paramètres associés sont généralement utilisés lors des études cliniques de phase II, lorsque l'objectif est de démontrer que le traitement est fonctionnel, au sens où il améliore la fonction de CFTR en tant que canal (indépendamment d'un éventuel bénéfice clinique).

La mesure de la concentration en ions chlorure dans la sueur est la plus fréquemment utilisée pour évaluer l'effet du traitement sur la fonction de CFTR. En effet, dans les glandes sudoripares, l'absence de canal ions chlorure médié par CFTR limite la capacité des cellules à réabsorber le chlorure et donc le sel, menant à la production d'une sueur plus salée que la normale (Quinton 1983). Cette mesure permet d'ailleurs d'établir le diagnostic de la mucoviscidose chez les nouveaux-nés, comme évoqué plus tôt dans ce manuscrit. Le rétablissement du canal CFTR par le traitement permet une réabsorption du surplus d'ions chlorure dans la sueur, réduisant sa teneur en sel. La mesure de ce paramètre n'est pas douloureuse, facile à réaliser, et peu coûteuse (Ramsey 2015).

L'amélioration des conditions de vie des patients peut aussi être évaluée par la réduction du nombre d'exacerbations pulmonaires, des événements menant à une chute rapide et conséquente des fonctions pulmonaires, de manière au moins partiellement irréversible. La résolution des exacerbations pulmonaires étant généralement associée à la prescriptions d'antibiotiques (Wagener 2013), administrés en intraveineuse ou par voie orale, une mesure simple de l'efficacité d'un traitement est donc le décompte de la quantité et de la durée d'antibiotiques prescrits sur une période donnée après initiation de celui-ci.

3.2 Historique du développement des traitements par modulateurs

Au cours des années 2000 débute la recherche de molécules pouvant moduler la stabilité et l'activité de CFTR par criblage haut débit (en anglais High throughput screening, HTS). Galletta et al. proposent un test basé sur la fluorescence dès 2001 (Galletta 2001). Des ions iodures sont ajoutés au niveau de la membrane des cellules, et leur quantité est mesurée par une protéine sonde, une yellow fluorescente protéine, sensible à ces ions. S'il est présent à la membrane, le canal CFTR est maintenu ouvert par l'ajout de forskoline, une protéine qui augmente l'AMP cyclique et donc favorise l'ouverture du canal (Hwang 1999). La cinétique de diminution de la fluorescence correspond à la capacité d'absorption des ions chlorure par la cellule via le canal CFTR et donc à l'activité du canal.

D'autres tests ont été proposés, tous visant à la mesure de l'activité du canal CFTR (Pedemonte 2005, Norez 2006, Carlile 2007)

Le HTS consiste à robotiser ces tests afin de pouvoir tester plusieurs milliers de molécules par jour. Celles qui montrent un effet positif pour l'ouverture du canal sont conservées et potentiellement modifiées chimiquement afin d'optimiser leur efficacité. Au début des années 2010, l'entreprise Vertex identifie ainsi plusieurs molécules d'intérêt pour la mucoviscidose, dont 4 ont fait l'objet d'études cliniques poussées.

3.2.1 IVACAFITOR

Vertex souhaite tout d'abord développer des potentiateurs, c'est-à-dire des molécules permettant d'améliorer la fonction de la protéine CFTR déjà présente à la membrane. 228000 molécules ont été criblées par HTS, et les structures chimiques des touches ont ensuite été optimisées, menant à l'identification du composé VX-770 (Ivacaftor) (figure 9) (Van Goor 2009). Ce composé était particulièrement intéressant en raison de sa bonne spécificité et de sa capacité à potentialiser plusieurs formes de CFTR. En particulier, le VX-770 augmente la sécrétion d'ions chlorures médiée par CFTR, dans des cellules primaires et dans des lignées cellulaires, pour les deux mutations les plus courantes que sont F508del et G551D. L'effet observé était plus important pour les mutations de classe III (dont fait partie G551D) que pour celles de classe II (dont fait partie F508del).

La molécule est entrée en phase d'essais cliniques pour les patients porteurs de la mutation G551D sur au moins un allèle, dont les premiers résultats ont été publiés en 2010. L'étude portait

sur une quarantaine de patients, et a montré une diminution significative de la concentration en ions chlorure dans la sueur des patients. De plus, bien que l'amélioration n'était pas significative, une tendance à l'amélioration des fonctions respiratoires fut observée (Accurso 2010). Cette étude fut ensuite complétée par une seconde portant sur des patients aux mêmes caractéristiques (Ramsey 2011), une troisième portant sur des enfants de 6 à 11 ans porteurs de la mutation G551D (Davies 2013), et par une quatrième sur des enfants (de plus de 6 ans) et adultes porteurs d'autres mutations de classe III. Dans les trois cas, il fut mis en évidence une baisse de la concentration en ions chlorure dans la sueur ainsi qu'une amélioration significative de la fonction respiratoire, mesurée par l'évolution du VEMS (voir section sur les paramètres cliniques mesurés). Enfin, l'efficacité du VX-770 (combinée au VX-809, voir ci-après) a été montrée pour les patients porteurs d'au moins une mutation de classe IV V ou VI (Rowe 2017). Néanmoins, l'étude sur les patients homozygotes pour la mutation F508del n'a pas permis de mettre en évidence une amélioration significative chez ces patients (Flume 2012).

De manière générale, les études cliniques sur VX-770 ont aussi permis de mettre en évidence un effet bénéfique de la molécule pour le gain d'IMC, et pour la réduction des infections, conséquences potentielles de l'amélioration de la fonction respiratoire. Néanmoins, pour toutes les études citées, c'est toujours le VEMS qui est utilisé comme donnée de référence pour mesurer l'efficacité du traitement.

L'utilisation du VX-770 dans le traitement de la mucoviscidose est autorisée depuis 2012 par la FDA aux Etats-Unis et par l'EMA en Europe. Initialement pour les adultes, l'âge de début de traitement a progressivement baissé pour être de 4 mois aujourd'hui, grâce aux études cliniques menées chez des patients de moins de 5 ans (Davies 2016) puis de moins d'un an (Davies 2021), montrant une diminution de la concentration en ions chlorure dans la sueur, mais aussi une amélioration des fonctions pancréatiques. Des fonctions anormales du foie ont été retrouvées chez 30% des enfants traités, mais ces effets sont réversibles (Davies 2016). L'utilisation de la molécule est autorisée pour une liste de mutations de classe III et supérieure, contenant G551D. En 2021, la FDA l'autorise pour 23 mutations de plus que l'EMA.

Depuis la mise sur le marché de la molécule, de nouvelles études ont été réalisées pour étudier l'évolution d'autres paramètres cliniques d'intérêt. Ont ainsi été démontrées une baisse significative du niveau de colonisation par *Pseudomonas aeruginosa* (Heltsche 2012), par *Staphylococcus aureus* et par *Aspergillus* (Frost 2019); une amélioration significative de l'IMC (Borowitz 2016); ou encore une amélioration des conditions de vie telles que ressenties par le patient (Quittner 2015), chez des patients porteurs de mutations de classe III ou supérieure.

La plupart des études cliniques citées sont réalisées sur 24 semaines. Des études de suivi sur de plus longues durées sont désormais réalisées pour mesurer l'effet de la molécule à plus long terme. De telles études ont été réalisées parallèlement au Canada (Kawala 2021), au Royaume-Uni (Mitchell 2021), en France (Hubert 2021) et aux USA (Guimbellot 2021) sur des patients ayant débuté le traitement par VX-770 depuis plus de 2 ans. Ces études ont montré que le traitement continuait à être bien toléré à long terme, et que les patients traités avaient moins d'infections et un déclin plus lent de leurs fonctions respiratoires. En conséquence, les patients traités ont un meilleur taux de survie et ont moins souvent besoin d'être transplantés. Cependant, l'évolution de la concentration en ions chlorure dans la sueur apparaissait comme non corrélée à la fonction respiratoire ou à l'IMC, posant la question de l'utilité de cette mesure pour l'évaluation de l'efficacité du traitement, et de la nature précise des mécanismes biologiques responsables de l'amélioration respiratoire.

En effet, le mécanisme d'action de VX770 n'est que partiellement connu. La potentialisation de CFTR par la molécule serait dépendante de l'état de phosphorylation du canal, mais indépendante de l'ATP puisque qu'elle est efficace sur des protéines dont le site NBD1 est muté, et dans d'autres où le site NBD2 est absent (Mutyam 2017).

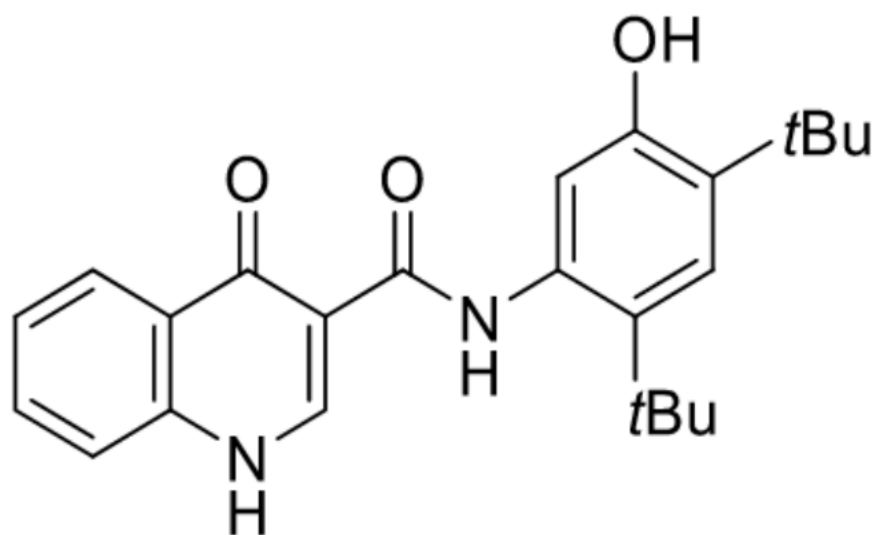


Figure 9 : Formule chimique du VX-770 (ivacaftor) (Bardin 2021)

3.2.2 LUMACAFITOR

En parallèle du développement de potentiateurs, Vertex a aussi souhaité développer des correcteurs, i.e. de petites molécules qui amélioreraient la maturation de CFTR jusqu'à son insertion dans la membrane, en concentrant ses efforts sur la correction de la mutation F508del. De la même manière que pour la caractérisation du VX-770, la méthode de détection de ces molécules incluait un HTS, réalisé sur 164000 molécules (Van Goor 2006). Cependant, contrairement au cas du VX-770, les cellules ont été préincubées avec les molécules pendant 48h avant de réaliser la mesure de l'activité du canal CFTR, pour laisser suffisamment de temps à CFTR pour être synthétisé et transporté à la membrane.

Les molécules étaient sélectionnées pour leur capacité à améliorer le processing de la protéine CFTR F508del dans le RE et leur transfert jusqu'à la membrane cellulaire. Pour cela, la quantité de protéine quittant l'appareil de Golgi était mesurée par immunoblotting, en tirant avantage de l'augmentation de la masse moléculaire de la protéine après sa glycolysation (de 135kDa à 180 kDa) (Cheng 1990). La quantité de protéine CFTR dans la cellule pouvait donc être mesurée dans ces deux états.

Plusieurs molécules d'intérêt ont été identifiées à partir de cette étude. Cependant, cette dernière avait été réalisée sur une lignée de thyroïde de rat, et l'efficacité ainsi que la cytotoxicité des molécules identifiées était très variable en fonction du type cellulaire étudié (Pedemonte 2010).

Une optimisation de la structure chimique des molécules touchées a donc été réalisée, menant à la mise au point de VX-809 (Lumacaftor) (figure 10). L'administration de cette molécule dans une lignée épithéliale humaine exprimant la protéine CFTR F508del (Human Epithelial Kidney 293, HEK-293) a entraîné la multiplication par 6 de la quantité de CFTR F508del qui quittait le RE (Van Goor 2006). Pour vérifier l'hypothèse d'une augmentation de la stabilité de la protéine CFTR F508del, une mesure du rythme de dégradation de la protéine après blocage de l'export des protéines du RE vers le Golgi par la baflomycine A (Takatsuki 2002) a été réalisée. Le rythme de dégradation était plus lent en présence de VX-809, ce qui témoigne en faveur d'un rôle positif du VX-809 sur la stabilisation de la protéine dans le RE. Enfin, une vérification de l'amélioration de l'activité canal ions chlorure liée à CFTR fut réalisée sur des cultures de cellules primaires de l'épithélium de bronches de patients homozygotes pour la mutation F508del. Après 48h d'incubation en présence de VX-809, les cellules traitées multipliaient par 8 la maturation de CFTR et par 4 l'activité du canal ions chlorure par rapport aux cellules non traitées. Tous ces résultats ont mené au lancement d'études cliniques sur des patients porteurs de mutations de classe II ou supérieure.

La tolérance et l'efficacité du VX-809 ont été évaluées au cours d'une étude clinique multicentrique chez des adultes homozygotes pour la mutation F508del (Clancy 2012). Une partie des patients a reçu un placebo afin de pouvoir mesurer l'effet de VX-809 seul. Si l'étude a conclu à la bonne tolérance du traitement chez les patients, aucune amélioration significative des fonctions respiratoires n'a pu être rapportée (évaluée par le VEMS). De même, aucune augmentation de l'activité du canal ions chlorure de CFTR dans l'épithélium nasal, mesurée par différence de potentiel nasal (Rowe 2011), n'a pas pu être mise en évidence. Pourtant, la concentration en ions chlorure dans la sueur était significativement réduite par le traitement, ce qui suggère que l'évolution de ce paramètre n'est pas corrélée à une amélioration clinique chez le patient.

Malgré une augmentation de la quantité de protéines qui migrent à la membrane, VX-809 seul ne semble ainsi pas suffire à rétablir la fonction de CFTR muté. Vertex a donc essayé d'associer ce correcteur avec un potentiateur de la fonction de CFTR, afin d'étudier si la combinaison de ces deux molécules permettait de rétablir un fonctionnement normal de CFTR (voir section sur les combothérapies)

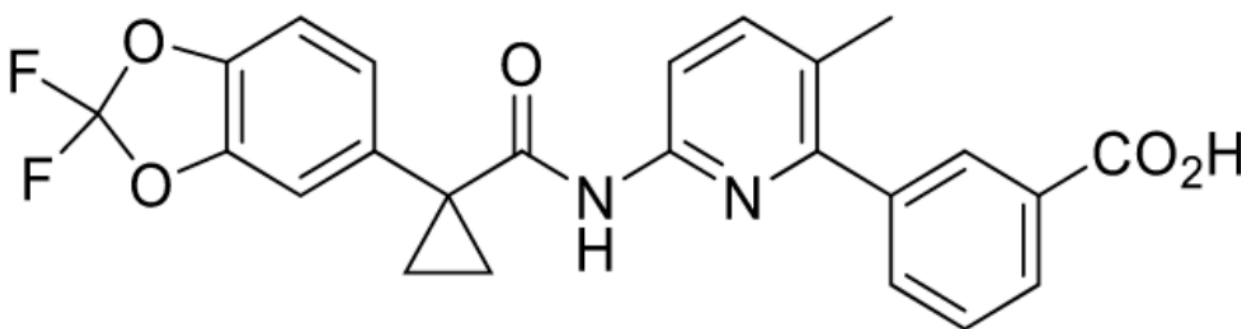


Figure 10 : Formule chimique du VX-809 (lumacaftor) (Bardin 2021)

Le mécanisme d'action exact du VX-809 reste incompris. Plusieurs études ont été menées pour tenter de comprendre ce mécanisme, et pointent vers des effets divers. VX-809 agirait sur la stabilisation du repliement de la protéine en modifiant la conformation du site MSD1 (Ren 2013), en se liant au site NBD1 (Hudson 2017), ou en améliorant les interactions entre les sites MSD1, MSD2 et NBD1 (Laselva 2018). Il a par ailleurs été montré que l'activité de VX-809 n'est pas restreinte à CFTR, cette molécule pouvant corriger d'autres protéines mutées de la même famille que CFTR, notamment ABCB4, une protéine impliquée dans le transport de la bile (Liu 2019). Il est

donc possible que VX-809 interagisse avec d'autres protéines, menant à des effets sur la physiopathologie de la mucoviscidose indépendamment de la correction de CFTR.

3.2.3 TÉZACAFTOR

Vertex a constaté que le traitement par VX-809 dans les cellules épithéliales humaines entraînait l'activation de l'enzyme CYP3A, qui métabolise les petites molécules étrangères à l'organisme telles que le VX-809 (Schneider 2018). De même, la molécule semble interagir facilement avec d'autres médicaments, et notamment interférer avec VX-770, limitant l'effet résultant de leur interaction (les données associées ont été conservées par l'entreprise et ne sont pas publiées). En prenant ces limitations en considération, l'entreprise a synthétisé le VX-661 (Tezacaftor) (figure 11), un correcteur de seconde génération dont la structure est très similaire à celle du VX-809. Contrairement à ce dernier, il n'active pas les enzymes CYP3A4, et peut donc théoriquement rester actif plus longtemps dans le RE (Donaldson 2018) .

Dès la première étude clinique, VX-661 a été testé seul et en association avec un correcteur, VX-770 (Donaldson 2018), chez des patient homozygotes pour la mutation F508del et chez des patients hétérozygotes F508del/G551D. S'agissant d'une étude de phase II, l'objectif était de montrer que le traitement agissait effectivement sur la fonction de CFTR. La mesure de référence pour cette étude a donc été l'évolution de la concentration en ions chlorure dans la sueur des patients, qui est liée à l'ouverture du canal ions chlorure de CFTR (Rowe 2007). Mais l'intérêt de cette mesure reste débattu : le cas du VX-770 avait déjà montré que la concentration en ions chlorure dans la sueur n'est pas nécessairement corrélée à l'évolution des fonctions respiratoires et donc à l'amélioration des conditions de vie des patients (Fidler 2017). L'évolution du VEMS était aussi mesurée dans l'étude, mais prise en compte secondairement.

Plusieurs concentrations de molécule ont été testées. Une baisse significative de la concentration en ions chlorure dans la sueur ainsi qu'une hausse significative du VEMS ont été mis en évidence pour les concentrations les plus élevées de VX-661 seul et pour la combinaison VX-661/VX-770 chez les patients F508del homozygotes, suggérant un effet bénéfique de VX-661. La meilleure amélioration pour ces deux paramètres a été obtenue par la combinaison VX-661/VX-770.

Les patients hétérozygotes F508del/G551D étaient déjà sous traitement par VX-770 lors de l'étude, et ce traitement n'a pas été interrompu. C'est donc la combinaison VX-661/VX-770 qui a été testée chez ces patients, montrant des améliorations significatives dans les deux paramètres étudiés.

A la lumière de ces résultats, les efforts pour le développement de traitements impliquant VX-661 se sont concentrés sur les combothérapies (voir section suivante).

VX-661 ayant été synthétisé en optimisant le VX-809 (Donaldson 2018), son mode d'action sur CFTR est supposé similaire à celui de cette dernière, et n'est donc pas totalement connu.

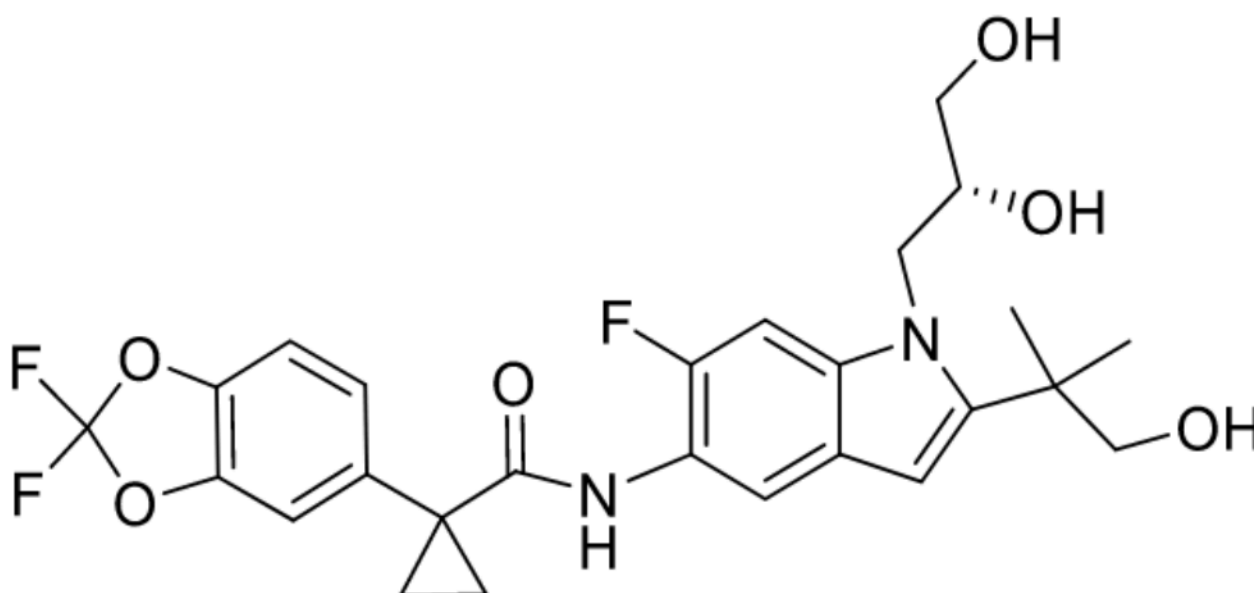


Figure 11 : Formule chimique du VX-661 (tezacaftor) (Bardin 2021)

3.2.4 ELEXACFTOR

Pour le développement d'une 3ème génération de correcteurs, Vertex a recherché des molécules qui auraient un effet positif sur le repliement de CFTR et qui agiraient de manière synergique avec d'autres correcteurs déjà existants. Un HTS a donc été réalisé en effectuant les tests *in vitro* sur des cellules bronchiques humaines, avec les molécules testées seules d'une part, et en combinaison avec VX-661 et VX-770 d'autre part pour tester un effet synergique. Deux molécules d'intérêt ont été retenues à partir de ce screening : le VX-659 (bamocaftor) (Davies 2018) et le VX-445 (elexacaftor) (Keating 2018), dont la formule chimique est représentée sur la figure 12.

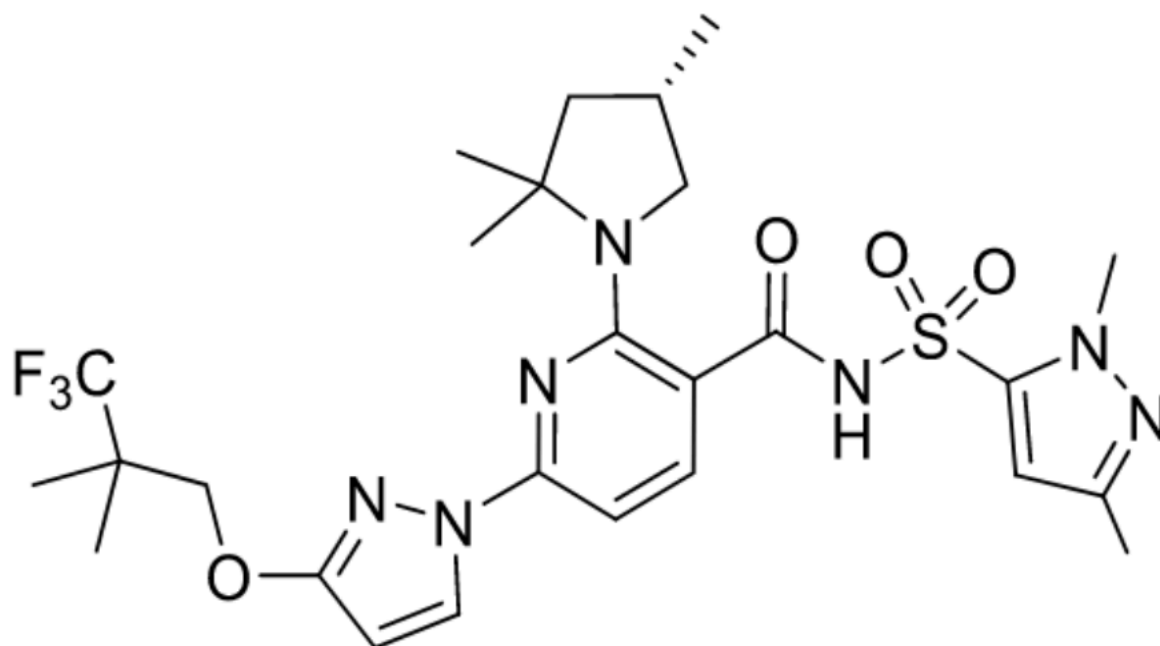


Figure 12 : Formule chimique du VX-445 (Elexacaftor) (Bardin 2021)

3.3 L'essor des combothérapies

VX-770, en tant que potentiateur, est le seul modulateur de Vertex ayant été utilisé en clinique sans être associé à d'autres molécules. Il est commercialisé depuis 2012 en Europe sous le nom Kalydeco (site de l'EMA). Cependant, comme expliqué dans la section précédente, son utilisation n'est autorisée que pour des mutations de classe III et supérieure.

Pour les mutations de classe II, dont F508del, la plus fréquente dans la mucoviscidose, les études se sont axées sur le développement de combothérapies impliquant simultanément un ou plusieurs correcteurs du repliement de CFTR et un potentiateur de l'activité de la protéine à la membrane.

3.3.1 LUMACAFITOR/IVACAFITOR (ORKAMBI)

Une première étude de phase II fut menée pour la combinaison VX-809/VX-661 sur des patients de plus de 18 ans homozygotes et hétérozygotes F508del, basée sur la mesure de l'évolution du VEMS et de la concentration en ions chlorure dans la sueur (Boyle 2014). Si l'amélioration du VEMS n'était pas significative pour les patients hétérozygotes, l'étude a mis en évidence une baisse significative de la concentration en ions chlorure dans la sueur ainsi qu'une hausse légère mais significative du VEMS chez les patients homozygotes traités. Ces résultats

encourageants ont justifié la réalisation d'une étude de phase III sur les patients homozygotes F508del (Wainwright 2015). Dans cette étude portant sur des patients de 12 ans et plus, le traitement était associé à une hausse significative du VEMS (donc des fonctions respiratoires) et de l'IMC, ainsi qu'à une baisse des infections pulmonaires et des hospitalisations.

Ces résultats ont mené à l'acceptation pour mise sur le marché de la combinaison VX-770/VX-809 par la FDA en juillet 2015 et par l'EMA en novembre 2015, ces deux entités autorisant l'utilisation de cette combothérapie pour le traitement des patients de 12 ans et plus homozygotes pour la mutation F508del.

Les effets du traitement à plus long terme ont été mesurés sur les patients de l'étude de phase III (Konstan 2017). Alors que le suivi initial était de 6 mois, ce nouveau suivi a été réalisé jusqu'à deux ans après le début du traitement. Il a mis en évidence le retour d'un déclin des fonctions respiratoires (mesurées par le VEMS). Cependant ce déclin s'effectuait à une vitesse 40% plus lente que chez les patients non traités. De plus, les patients traités continuaient à améliorer leur IMC, et étaient moins sujets aux infections et aux hospitalisation que ceux sans traitement. Ainsi, la combinaison VX-770/VX-809 apparaissait comme étant toujours bénéfique aux patients après deux ans.

Deux autres études ont ensuite été menées pour évaluer l'effet de cette combothérapie sur les patients plus jeunes, âgés de 6 à 11 ans (Ratjen 2017, Milla 2017). Elles n'ont pas permis de mettre en évidence une amélioration significative du VEMS, ce qui peut en partie s'expliquer par une valeur déjà proche de la normale en moyenne chez les patients avant le début du traitement. En revanche, elles ont montré une baisse significative de la concentration en ions chlorure, une hausse de l'IMC, et une hausse du LCI (une autre méthode de mesure des fonctions pulmonaires, voir section sur les paramètres cliniques). Ces données ont mené à l'élargissement de l'autorisation de traitement aux patients de 6 ans et plus en 2018, puis de 2 ans et plus en 2019 (Roda 2022).

L'évolution de la concentration en ions chlorure dans la sueur lors du traitement par la combinaison VX-770/VX-809 apparaît comme n'étant corrélée ni aux fonctions pulmonaires, ni à l'IMC (Sagel 2021, Bui 2021).

Dans chacune des études décrites précédemment, les patients dont les fonctions respiratoires étaient très faibles avant de débiter le traitement étaient exclus. Leur cas fut considéré dans une étude distincte, montrant que pour ces patients, le traitement restait sûr si les doses étaient réduites, et qu'il menait à une réduction de la durée de recours aux antibiotiques et du nombre d'hospitalisations. En revanche, le VEMS ne s'améliorait pas chez ces patients. Ces données suggèrent que les patients répondent différemment au traitement suivant la valeur de leurs

paramètres cliniques à l'initiation du traitement. Les résultats des études de phase III semblent aller dans le même sens : les effets du traitement ont été évalués séparément sur différents sous-groupes définis par leur niveau de fonction pulmonaire au début du traitement (c'est-à-dire séparés par tranche de valeur de VEMS). Si tous les sous-groupes ont effectivement connu une amélioration significative de ces fonctions, leur niveau d'évolution n'était pas le même : les patients aux fonctions les plus détériorées connaissaient une amélioration moins importante de leur VEMS (Elborn 2016).

Certains effets bénéfiques non liés aux fonctions respiratoires peuvent aussi apparaître chez certains patients pour lesquels le traitement n'améliore pas le VEMS. Il a notamment été montré que les patients traités recouraient moins aux antibiotiques même si leur VEMS n'augmentait pas (McColley 2019). Par ailleurs, la valeur de VEMS au début du traitement joue sur la capacité de ce dernier à l'améliorer. Lorsque la valeur de départ est proche de la normale, on ne mesure pas d'amélioration significative des fonctions pulmonaires (Sagel 2021).

On note cependant que pour les patients homozygotes F508del pour lesquels le traitement mène à une augmentation du VEMS, cette dernière reste moins importante que chez les patients pouvant être traités par le VX-770 seul. Il apparaît en réalité que le VX-770 limite l'effet correcteur du VX-809 (Cholon 2014). De plus, l'activation du CYP3A4 par le VX-809 limite l'effet du traitement (Schneider 2018).

3.3.2 TEZACAFTOR/IVACAFTOR (SYMKEVI)

Face aux effets certes bénéfiques mais limités de la combinaison VX-770/VX-809, le correcteur VX-809 a été remplacé par le VX-661, qui n'interagit pas avec le VX-770, et n'active pas CYP3A4 (Schneider 2018).

Les essais cliniques sur des patients de 12 ans et plus, homozygotes pour la mutation F508del, ont montré des effets bénéfiques au niveau des fonctions pulmonaires (mesuré par le VEMS) et au niveau de la durée de recours aux antibiotiques (Taylor-Cousar 2017), mais ces améliorations n'étaient pas supérieures à celles observées chez les patients traités par la combinaison VX-770/VX-809.

En revanche, contrairement à la combinaison VX-770/VX-809, des effets bénéfiques ont aussi été mesurés chez les patients de 12 ans et plus hétérozygotes possédant un allèle F508del et un allèle correspondant à une mutation de la classe III ou supérieure. En particulier, une baisse significative de la concentration en ions chlorure dans la sueur a été rapportée, ainsi qu'une amélioration significative du VEMS, supérieure à celle atteinte avec le VX-770 seul (Rowe 2017).

Basé sur ces données, le traitement a été autorisé par la FDA et par l'EMA en 2018, pour les patients de 12 ans et plus hétérozygotes possédant un allèle F508del et un allèle correspondant à une mutation de classe III ou supérieure. L'autorisation a rapidement été élargie aux patients de plus de 6 ans après la parution d'une étude montrant l'efficacité du traitement à cet âge via la diminution de la concentration en ions chlorure dans la sueur et malgré l'absence d'amélioration en VEMS (Walker 2019). Elle sera complétée par la suite par une seconde étude (Davies 2021) utilisant le LCI comme mesure de la fonction pulmonaire et montrant une amélioration significative de ce paramètre clinique.

De manière intéressante, l'environnement des patients peut aussi jouer un rôle sur l'effet des traitements. Ainsi, la combinaison VX-661/VX-770 est bien moins efficace chez les patients exposés aux fumées de cigarettes (Baker 2021).

En 2021, des études cliniques étaient toujours en cours pour montrer l'effet du traitement chez les patients de moins de 6 ans, et ainsi étendre l'autorisation de mise sur le marché de la combinaison VX-661/VX-770 à ces âges.

3.3.3 TRIPLE COMBINAISON VX-770/VX-661/VX-445 (TRIKAFTA)

Une avancée thérapeutique majeure a eu lieu dans la mucoviscidose avec le développement de la trithérapie VX-770/VX-661/VX-445. Elle est très récente puisque l'autorisation de mise sur le marché a été donnée en 2019 par la FDA et en 2020 par l'EMA.

Dans un premier temps, le VX-659 et le VX-445 ont été retenus pour être testés en combinaison avec les VX-770 et VX-445. Deux études cliniques de phase II ont donc été réalisées en parallèle pour mesurer l'effet des combinaisons VX-770/VX-661/VX-659 (Davies 2018) et VX-770/VX-661/VX-445 (Keating 2018), sur des patients homozygotes pour la mutation F508del et sur des patients hétérozygotes F508del porteurs d'une mutation de classe I (pour lesquels aucun traitement n'existait alors). Les résultats ont été impressionnants, avec notamment une baisse significative de la concentration en ions chlorure dans la sueur dans les deux cas, mais aussi et surtout une amélioration très importante des fonctions respiratoires, mesurée par le VEMS, à un niveau plus important que toutes les autres thérapies évoquées précédemment. De manière encore plus impressionnante, sur les patients homozygotes pour la mutation F508del qui recevaient déjà la combinaison VX-770/VX-661 avant de tester la triple combinaison, cette dernière a permis d'encore augmenter de manière significative les fonctions respiratoires.

Seul le VX-445 a été considéré pour les études suivantes, celui-ci ayant montré un effet légèrement supérieur au VX-659. Les résultats démontrés en phase II ont très vite été confirmés,

d'une part pour les patients homozygotes pour la mutation F508del (Heijerman 2019), et d'autre part pour les patients hétérozygotes F508del porteurs d'une mutation de classe I (Middleton 2019), âgés de 12 ans et plus. Dans les deux cas, le traitement était associé à une augmentation très importante du VEMS, une amélioration de la qualité de vie des patients telle que ressentie par ces derniers (mesurée par un questionnaire), une baisse du recours aux antibiotiques et une baisse de la concentration en ions chlorure dans la sueur. Dans le cas des patients hétérozygotes, l'effet du traitement était mesuré en comparant des patients traités et non traités. En revanche, au moment de l'étude, tous les patients homozygotes F508del étaient déjà traités par une combinaison de modulateurs (VX-770/VX-809 ou VX-661/VX-770). Pour ce cas, l'effet a donc été mesuré en comparant des patients traités avec VX-661/VX-770 à d'autres qui arrêtaient le traitement VX-661/VX-770 pour débiter celui basé sur la triple combinaison.

A la lumière de ces résultats, la triple combinaison VX-770/VX-661/VX-445 a été autorisée par la FDA fin 2019, et par l'EMA en 2020, pour les patients de 12 ans et plus homozygotes ou hétérozygotes pour la mutation F508del (Roda 2022)

Récemment, une dernière étude a été réalisée sur des patients homozygotes ou hétérozygotes F508del âgés de 6-11 ans pour évaluer l'effet de la triple combinaison dans cette tranche d'âge (Zemanick 2021). Etant donné qu'une partie des patients étudiés pourraient être traités avec d'autres combinaisons (les homozygotes F508del notamment), il n'y a pas eu de comparaison à un groupe non traité dans cette étude, et l'évolution des différents paramètres étudiés a donc été mesurée de manière absolue. Les résultats montrent que le traitement est sûr, ainsi qu'une amélioration significative du VEMS, du LCI, de la concentration en ions chlorure dans la sueur et de l'IMC. Ils pourraient servir de base pour un élargissement de l'autorisation de mise sur le marché de la triple combinaison pour les patients de moins de 12 ans.

3.3.4 DÉVELOPPEMENT D'AUTRES MODULATEURS

Vertex développe d'autres potentioteurs, avec l'objectif que ces derniers soient plus stables que le VX-770, ce qui permettrait d'augmenter l'espacement entre deux prises du médicament chez le patient. Vertex a ainsi développé le VX-561, actuellement à l'étude quand combiné aux VX-445 et VX-661, ou au VX-659 et VX-661 (Bardin 2021). En parallèle, l'entreprise continue de rechercher de nouveaux correcteurs.

Une autre entreprise, Galapagos NV, a lancé sa propre pipeline de recherche de correcteurs et potentioteurs de CFTR. Ont ainsi été notamment développés les potentioteurs ABBV-974, ABBV-2451, et ABBV-3067, et les correcteurs ABBV-2222, ABBV-2737 et ABBV-3221. L'effet de

ces différents modulateurs utilisés seuls ou en association est actuellement l'objet d'études cliniques. Des premiers résultats ont déjà montré un effet positif de l'ABBV-974 chez des patients avec une mutation de classe III, à un niveau similaire à celui du VX-770 (Davies 2019). De même, l'utilisation de VX-770 avec l'ABBV-2222 sur des patients homozygotes pour la mutation F508del montre un effet limité mais positif sur les fonctions pulmonaires (Bell 2019). Enfin, l'administration d'ABBV-2737 chez des patients homozygotes F508del déjà traités par la combinaison VX-770/VX-809 a permis d'augmenter encore la réduction en concentration d'ions chlorure dans la sueur, et d'améliorer légèrement leurs fonctions pulmonaires (van Koningsbruggen-Rietschel 2020).

Néanmoins, si de nombreuses recherches sont en cours dans le domaine des modulateurs de CFTR, la triple combinaison VX-770/VX-661/VX-445 reste aujourd'hui la référence des traitements contre la mucoviscidose. Aujourd'hui, Vertex et AbbVie sont les deux entreprises majeures poursuivant la recherche dans ce domaine.

3.4 Nouveaux défis pour la détection de bénéfices cliniques

Comme on a pu le voir dans les sections précédentes, le processus de mise sur le marché pour les modulateurs est toujours sensiblement le même. Il débute par une vérification de l'effet bénéfique de la ou des molécules d'intérêt sur l'augmentation de l'activité de CFTR dans des cultures cellulaires (lignées puis primaires). Une étude clinique de phase II est ensuite menée sur des patients adultes (ou au minimum de plus de 12 ans) en évaluant l'évolution de la concentration en ions chlorure dans la sueur et du VEMS. La mesure de ce second paramètre a par ailleurs progressivement pris le pas sur celle de la concentration en ions chlorure de la sueur à mesure que des études démontraient l'absence de corrélation entre cette valeur et la qualité de vie des patients, notamment dans le cas du VX-770 (Durmowicz 2013, Barry 2014), mais aussi pour VX-770/VX-809 (Masson 2019). Ainsi, pour les études qui ont mené à la mise sur le marché de VX-661/VX-770 et de VX-770/VX-661/VX-445, la valeur du test de la sueur était toujours mesurée, mais n'était plus le paramètre clinique d'intérêt principal (Taylor-Cousar 2017, Heijerman 2019). Dans un troisième temps, une étude de phase III est réalisée, prenant systématiquement le VEMS comme paramètres clinique d'intérêt principal. La mesure d'une amélioration significative pour ce paramètre (combinée à la confirmation de la sûreté du produit pour le patient) mène à la délivrance d'une autorisation de mise sur le marché par la FDA et l'EMA.

De nouvelles études sont alors réalisées pour élargir l'autorisation de mise sur le marché à des patients plus jeunes, jusqu'à 6 ans dans un premier temps. Cependant les patients de ces âges ayant un VEMS plus proche de la normale, il devient plus délicat de démontrer une amélioration significative dans ce paramètre. Ce fut notamment le cas pour les études sur les combinaisons VX-770/VX-809 (Ratjen 2017) et VX-770/VX-661 (Ratjen 2017), pour lesquelles le VEMS n'augmentait pas. Ont été alors étudiées les évolutions de la concentration en ions chlorure dans la sueur des patients, de l'IMC, et du LCI.

Enfin, dans certains cas, les études sont élargies aux enfants de moins de 6 ans. Ce fut notamment le cas pour le VX-770 seul (Davies 2016). Le VEMS étant normal chez ces enfants, d'autres mesures sont prises en référence, notamment le gain d'IMC, le LCI, et la baisse de la concentration en ions chlorure dans la sueur.

Ainsi, si le VEMS reste la mesure de référence pour étudier l'effet d'un traitement, il semble aujourd'hui atteindre ses limites. D'une part, il n'est pas utile dans le cas d'essais cliniques sur de jeunes enfants. D'autre part, les avancées spectaculaires permises par la triple combinaison VX-770/VX-661/VX-445 notamment retardent le déclin des fonctions respiratoires chez les patients, rendant difficile la mise en évidence de gain supérieurs par de potentiels traitements futurs, ou imposant de travailler avec une cohorte initiale de taille bien plus conséquente (Stanojevic 2016). La définition de nouveaux paramètres d'intérêt apparaît donc nécessaire pour pouvoir continuer à améliorer les traitements dans la mucoviscidose.

Ces problématiques sont évoquées depuis 2012 et le début de l'élargissement des essais cliniques du VX-770 sur les enfants de moins de 6 ans (Davies 2013). Le constat initial est que bien que les fonctions respiratoires restent normales durant les premières années de vie des patients, des symptômes sont décelables dès les premiers mois de vie. Des infections pulmonaires par bactéries sont décelables dès le deuxième mois chez certains enfants (Stafler 2011), et des paramètres inflammatoires (neutrophile elastase, interleukines notamment) sont sur exprimés avant 6 mois (Khan 1995). Un traitement débuté suffisamment tôt pourrait permettre de traiter ces problèmes avant qu'ils n'évoluent vers des dommages irréversibles. Il est donc nécessaire de trouver de nouveaux paramètres cliniques d'intérêt qui pourraient évoluer significativement en réponse à un traitement dès les premiers mois de vie.

Les recherches se portent notamment sur la mesure de paramètres gastro-entérologiques (Bodewes 2016). Les dommages liés au pancréas débutent en effet *in utero* (Abu-El-Haija 2012) bien que leur origine reste inconnue, et la détérioration se poursuit durant l'enfance (O'Sullivan 2013). L'efficacité d'un traitement pourrait donc se mesurer dès la naissance sur le pancréas. Lors

des essais du VX-770 sur les patients de moins de 2 ans porteurs d'une mutation de classe III (Rosenfeld 2018), des marqueurs de la fonction pancréatique (elastase-1 et TRI) ont ainsi été mesurés dans les selles pour évaluer un potentiel effet bénéfique, et une amélioration significative a été mise en évidence.

Cependant, une amélioration pancréatique n'est pas nécessairement liée à un bienfait au niveau respiratoire, et une mesure efficace des dysfonctionnements pulmonaires reste nécessaire. En effet, bien que les patients restent asymptomatiques pendant leurs premières années de vie, la dégradation de leurs poumons débute dès la naissance par une obstruction progressive des petites voies respiratoires (Tiddens 2010). Plusieurs études ont mis en avant l'utilité du recours au scanner pour visualiser ces petits dysfonctionnements, en particulier les bronchectasies et l'obstruction des voies respiratoires. Les défauts peuvent être détectés chez des patients qui sont encore asymptomatiques (Stick 2009), et les évolutions des scores associés au scanner semblent être liés à des défauts pulmonaires différents de ceux mesurés par la spirométrie. En effet, chez des patients non traités par des modulateurs, le déclin des scores de scanner est plus rapide et plus fréquent que celui du VEMS (De Jong 2005). De plus, le scanner apparaît supérieur au VEMS pour la détection de détériorations pulmonaires à très jeune âge. De Jong et al. ont ainsi montré que chez des enfants le score de scanner diminuait en deux ans alors que leur VEMS restait stable, voire augmentait (de Jong 2004).

Malheureusement, la réalisation de scanners est coûteuse, l'opération est lourde pour le patient puisqu'il est nécessaire de procéder à une anesthésie générale (pour s'assurer d'une respiration artificielle constante et contrôlée), les radiations limitent les possibilités de réalisation de scanner à un par an au plus, l'analyse prend du temps et demande le recours à des experts, et les appareils permettant de les réaliser ne sont pas disponibles dans tous les CRCM (Stick 2013).

Le LCI s'est progressivement développé comme mesure alternative pour évaluer l'état des poumons chez les très jeunes patients (Stanojevic 2016). Mais ce dernier présente aussi des limitations. Si la variabilité associée à cette mesure est très faible chez les personnes saines, elle est au contraire très élevée chez les patients atteints de la mucoviscidose (Jensen 2013), et une taille de cohorte conséquente est donc nécessaire pour mesurer une amélioration significative de ce paramètre.

De plus cette variabilité dépend de la gravité de la maladie (Lum 2013), ce qui suggère de séparer les patients en plusieurs niveaux de sévérité avant d'utiliser ce paramètre dans une étude clinique. En particulier, le LCI serait plus intéressant pour des cohortes de jeunes patients dont les fonctions pulmonaires sont encore correctes que chez des patients plus âgés possédant des capacités respiratoires plus limitées.

Le LCI et le VEMS semblent aussi mettre en valeur différents aspects de la perte de fonctions respiratoires et de la réponse à un traitement. Ainsi, si le traitement d'exacerbations pulmonaires par des antibiotiques est associé à une hausse significative à la fois du VEMS et du LCI de manière globale, les réponses selon ces deux paramètres sont très hétérogènes et ne sont pas corrélées entre elles (Sonneveld 2015). De plus, le niveau de réponse aux antibiotiques dépend ici aussi de l'état du patient avant l'exacerbation pulmonaire, suggérant là encore de séparer les patients en plusieurs niveaux de sévérité au préalable.

La mesure de la concentration en ions chlorure dans la sueur présente elle aussi une limitation de taille : la variabilité inter patients est très grande selon ce paramètre, et son évolution n'est corrélée à aucun paramètre respiratoire (Fidler 2017). Elle ne semble donc pas permettre de refléter l'effet d'un traitement pour les patients.

3.5 La recherche d'un score de réponse aux traitements

Il apparaît de la section précédente qu'aucun paramètre clinique ne permet de refléter globalement la maladie et son évolution. Le fait de ne se concentrer que sur les gains en spirométrie présente le risque de passer à côté d'autres bénéfices pour le patient. Ce constat a suscité des réactions dès 2011 (Stick 2011), alors qu'un modulateur prometteur, le Denufosol, venait d'échouer en phase III, ne parvenant pas à démontrer une amélioration en VEMS ou en CVF (Ratjen 2012). Le constat était alors le suivant : l'administration de modulateurs pourrait être la plus efficace dès les premiers mois de vie, avant que les premières broncheectasies causant des dommages irréversibles aux poumons ne surviennent. Mais il est impossible de mesurer un effet à cet âge en prenant en compte seulement les résultats spirométriques. Des outils de visualisation des défauts pulmonaires alternatifs ont été proposés, comme le scanner ou le LCI, mais ces derniers ne sont pas parfaits, comme on a pu le noter précédemment. Partant de ce constat, des études ont tenté de développer un score unique, combinaison de plusieurs paramètres cliniques, permettant d'évaluer le mieux possible l'effet d'un traitement par modulateurs sur les conditions de vie des patients.

Le CF-ABLE score a été proposé en 2013 (McCarthy 2013). Il fut développé à partir de l'étude de l'évolution de plusieurs paramètres cliniques chez 49 patients atteints de la mucoviscidose entre 2004 et 2010, d'âge moyen 20 ans au début du suivi, dont 13 sont décédés ou ont été transplantés au cours du suivi. Les paramètres considérés étaient l'insuffisance pancréatique, la présence ou non d'un diabète, la colonisation par *Pseudomonas aeruginosa* ou *Staphylococcus*

aureus, la mutation de CFTR, le VEMS, l'IMC, la densité minérale osseuse, le nombre d'exacerbations, c'est-à-dire le nombre d'hospitalisations dues à une augmentation de l'intensité des symptômes, et l'âge. Les auteurs ont recherché la meilleure combinaison de ces paramètres qui permettrait de prédire un résultat négatif pour le patient, à savoir le décès ou une transplantation. Ils ont ainsi développé un score basé pour moitié sur le VEMS, et complété de manière équivalente par l'âge, l'IMC et le nombre d'exacerbations. Plus le score est élevé et plus les patients ont une forte probabilité d'avoir un résultat négatif (les patients avec un score de 5 ou plus ont 26% de chances d'avoir un tel résultat dans les 4 ans qui suivent la mesure).

En 2017, un autre score, basé sur l'étude du registre français de données cliniques de 8000 patients diagnostiqués avec la mucoviscidose entre 1992 et 2015 (Ellis 2015), a été proposé (Nkam 2017). L'objectif était similaire à celui du CF-ABLE score, à savoir associer à chaque patient une probabilité de décès ou de transplantation dans les trois prochaines années à partir de ses données cliniques. On retrouve aussi de nombreuses similitudes avec le CF-ABLE score, le VEMS comptant pour un tiers du score total, et l'IMC ainsi que le nombre d'exacerbations (mesuré par le nombre d'hospitalisations et de cures d'antibiotiques par an) sont aussi pris en compte. A ces mesures sont enfin ajoutés l'administration ou non de corticostéroïdes, et le niveau de colonisation par *Burkholderia cepacia*. Comme pour le CF-ABLE, plus le score est élevé et plus les patients ont de fortes probabilités d'avoir un résultat négatif (les patients avec un score de 5 ou plus ont 50% de chances d'avoir un tel résultat dans les 3 ans qui suivent la mesure). Ce score a ensuite été appliqué à des cohortes de patients des Etats-Unis (Ramos 2022) et du Royaume-Uni (Frost 2021). Malheureusement les bénéfices se sont avérés limités. La prédiction était bonne sur la cohorte américaine, mais pas supérieure à l'utilisation du VEMS seul. Sur la cohorte du Royaume-Uni, la prédiction était bien moins bonne que sur la cohorte française. Cela semble montrer que les paramètres à considérer diffèrent en fonction des cohortes, et qu'un score global ne pourrait être obtenu qu'à partir d'une étude prenant en compte des patients de plusieurs pays.

Les deux scores décrits précédemment sont construits en tant que combinaison de plusieurs paramètres cliniques. Lors de leur validation, ils ont montré être de meilleurs prédicteurs du décès ou de la transplantation des patients que chacune de leurs composantes considérées individuellement, ce qui semble confirmer l'aspect multidimensionnel des paramètres à prendre en compte pour caractériser l'évolution de la maladie et donc la nécessité de ne pas se concentrer sur un seul critère. Malheureusement, ces deux scores sont aussi difficilement utilisables dans le contexte de la mesure de l'effet des modulateurs. Ces derniers réduisant très fortement le nombre de

décès ou de transplantations des patients à court terme, un score évaluant la probabilité qu'un tel événement survienne ne semble pas adapté.

Plutôt que de prédire ces événements, une autre étude s'est concentrée sur la prédiction des exacerbations pulmonaires, qui sont associées à un déclin irréversible des conditions de vie des patients (Andrinopoulou 2020). Les paramètres considérés étaient le VEMS, l'IMC, la mutation de CFTR, l'insuffisance pancréatique, et la colonisation par *Staphylococcus aureus* ou *Pseudomonas aeruginosa*. Cependant aucune combinaison de ces paramètres ne s'est révélée plus efficace pour prédire les exacerbations pulmonaires que le VEMS seul.

Dans cette thèse nous avons privilégié une autre approche. Jusqu'ici, toutes les études se concentraient sur la recherche d'un score global, qui permettrait d'évaluer une réponse au niveau global pour une cohorte donnée. A l'inverse, nous avons recherché s'il était possible de déterminer des sous-cohortes de patients qui évolueraient de manière similaire selon certains paramètres cliniques. Pour cela, nous avons repris les paramètres les plus couramment mesurés pour évaluer l'effet d'un traitement par modulateur et évoqués précédemment (à l'exception du LCI dont l'utilisation n'est pas encore généralisée à tous les CRCM, au contraire des tests spirométriques), dans le but de mettre en lumière plusieurs profils de réponse associés à des combinaisons différentes de certains paramètres cliniques (première partie de résultats de cette thèse). Une seule autre étude s'est intéressée de manière similaire à l'hétérogénéité de la réponse à des traitements dans le cas de broncheectasies (Sibila 2021). S'intéressant aux évolutions comparées des valeurs de VEMS, du nombre d'exacerbations et du score de qualité de vie donné par les patients. Cette dernière a montré une absence de corrélation entre l'évolution de ces trois paramètres et surtout une très grande hétérogénéité dans la réponse à ces paramètres au sein de la cohorte étudiée, suggérant l'existence de plusieurs sous-groupes réagissant différemment au traitement et supportant donc notre approche.

4. Approches omiques et intérêt dans la mucoviscidose

Au cours des 20 dernières années, grâce notamment à l'amélioration de la puissance de calcul des ordinateurs, l'acquisition et l'analyse de données dites « omiques » s'est grandement simplifiée et a été de plus en plus utilisée pour étudier le fonctionnement de diverses maladies. Les données omiques donnent de précieuses informations sur les éléments qui permettent aux cellules étudiées de fonctionner.

L'analyse génomique décrit l'ADN présent dans les cellules, quand l'analyse transcriptomique étudie leur ARN (de différents types suivant l'analyse effectuée), l'analyse protéomique s'intéresse aux protéines et l'analyse métabolomique aux métabolites produits suite à la dégradation de protéines ou d'autres molécules. Chacune de ces méthodes ajoute un niveau de complexité et est complémentaire des autres.

L'utilisation de ces méthodes dans le cadre de la mucoviscidose pourrait permettre de mettre en lumière l'origine de certains symptômes et potentiellement de mener à la découverte de nouveaux traitements. En particulier, la protéomique et la transcriptomique apparaissent particulièrement intéressants pour comprendre les dysfonctionnements cellulaires induits par les mutations de CFTR, et les potentielles corrections apportées par les modulateurs.

4.1 L'analyse protéomique

L'étude de l'ADN et de l'ARN permet déjà une bonne caractérisation des cellules et des différences qui peuvent exister entre celles de différents individus, mais les modifications post traductionnelles complexifient encore cette diversité. Ainsi, à partir des 20000 gènes répertoriés chez l'homme par analyse génomiques (Lander 2001), on estime que l'on peut obtenir plusieurs millions de protéines et formes protéiques différentes (Aebersold 2018). L'épissage alternatif notamment, c'est-à-dire la possibilité de produire des ARNm différents à partir d'un même ADN, explique en partie le grand nombre de protéines qui peuvent être produites.

Le protéome (la liste des protéines) est complexe et varie en fonction des cellules. Les protéines agissent en synergie avec d'autres dans le but d'accomplir des fonctions biologiques, et des dysfonctionnements sur une protéine peuvent donc mener à une dérégulation de nombreuses autres, elles mêmes étant à l'origine de modifications dans le phénotype cellulaire.

Dans le cadre de la mucoviscidose, si l'on connaît le gène, CFTR, et les mutations à l'origine de la maladie, notre compréhension des phénotypes qui en résultent est encore très limitée.

Les études protéomiques ont ainsi pour but de mettre en lumière l'effet du dysfonctionnement de CFTR sur le fonctionnement de la cellule, mais aussi de mieux comprendre les différentes étapes menant à la maturation de CFTR et les protéines impliquées dans ce processus, ainsi que de déterminer quelles protéines interagissent avec CFTR lors de la réalisation de sa fonction canal ions chlorure. Mais l'analyse protéomique pourrait aussi s'avérer utile pour mieux comprendre l'effet des modulateurs, non seulement sur CFTR, mais aussi potentiellement sur d'autres protéines dérégulées par le dysfonctionnement de CFTR et dont la correction pourrait mener à des effets bénéfiques pour les patients. De telles protéines pourraient alors être ciblées par d'autres molécules plus spécifiques, et donc être à l'origine de nouveaux traitements dans la mucoviscidose.

La méthode la plus couramment utilisée aujourd'hui pour la réalisation d'analyses protéomiques est la spectrométrie de masse couplée à une chromatographie en phase liquide (LC-MS pour Liquid Chromatography-Mass Spectrometry) de lysats protéiques issus de la digestion des protéines par des protéases. C'est l'analyse dite bottom up (Zhang 2013).

Dans cette partie nous étudierons dans un premier temps le principe de fonctionnement de l'analyse bottom-up, puis nous verrons quelles avancées dans la compréhension de la maladie ont été permises par la protéomique, avant d'évoquer son utilité pour le futur de la recherche contre la mucoviscidose.

4.1.1 L'ANALYSE BOTTOM-UP

De nombreuses étapes sont nécessaires avant d'obtenir une liste de protéines exprimées dans un échantillon. Les cellules sont d'abord lysées, et les protéines extraites. Ces dernières sont ensuite digérées par des enzymes qui les coupent en peptides. Ceux-ci sont ensuite analysés et quantifiés par LC-MS. Enfin, les données de sortie du LC-MS sont analysées informatiquement.

Lyse cellulaire

Durant cette étape, la membrane des cellules est détruite pour que les protéines puissent être extraites. Ceci peut être réalisé mécaniquement ou chimiquement. Un exemple de lyse mécanique est l'utilisation d'un moulin à billes (Islam 2017), qui consiste à agiter rapidement de petites billes de verre avec les cellules, forçant l'ouverture des membranes lors des collisions entre billes et cellules (Chisti 1986). Des méthodes de sonication ont aussi été proposées, consistant à rompre les membranes cellulaires en utilisant des ultrasons (Marentis 2005). Les approches chimiques impliquent l'utilisation de détergent pour dissoudre les membranes cellulaires (Zhang 2017). Le

plus souvent, les deux méthodes, mécanique et chimique, sont utilisées conjointement pour améliorer l'efficacité de la lyse (Feist 2015).

Digestion enzymatique

Durant cette étape, les protéines sont séparées en peptides, plus petits, qui pourront être analysés en LC-MS. Cette étape est donc importante, une mauvaise séparation entraînant des erreurs de mesure lors des phases suivantes.

Cette digestion est généralement réalisée par la trypsine, une enzyme qui coupe les protéines au niveau des lysines et des arginines (Vandermarliere 2013). Cela permet d'obtenir des peptides de taille 500 à 3000Da, idéale pour l'analyse par LC-MS. Cependant l'utilisation de la trypsine reste imparfaite car une faible proportion des peptides générés sont trop longs ou trop courts pour l'analyse LC-MS. D'autres enzymes sont donc testées en complément de la trypsine pour ces cas particuliers (Giansanti 2016).

La digestion peut être réalisée en solution, en gel, ou sur filtre. La digestion en solution consiste en la dénaturation des protéines (désorganisation spatiale sans rupture des liaisons) qui facilite la digestion par la trypsine (ou une autre enzyme). Cette méthode présente l'avantage d'être rapide mais n'est pas la plus efficace en ce qui concerne le rendement.

La digestion en gel utilise, comme son nom l'indique, un gel permettant de séparer les protéines selon leurs masse moléculaire. Les protéines sont d'abord concentrées en haut du gel, puis chargées négativement, celles les plus lourdes devenant les plus chargées. Un courant est alors appliqué pour faire migrer les protéines, les plus lourdes se déplaçant le plus loin. Les protéines sont ensuite digérées par la trypsine sur le gel directement.

La digestion sur filtre consiste à utiliser des cartouches pour retenir les protéines et ainsi réaliser les étapes de lyse et de digestion successivement. Une de ces méthodes est le Suspension Trap (S-Trap), développé par Protifi (protifi.com). Les protéines sont dans un premier temps lysées et dénaturées, puis placées dans une cartouche où elles sont digérées plusieurs fois, ce qui permet une meilleure digestion avec une perte minimale de matériel grâce à la cartouche qui retient les protéines. Enfin, les peptides sont récupérés par centrifugation. Cette méthode est relativement récente (2014) mais de plus en plus utilisée.

Chromatographie en phase liquide et spectrométrie de masse

Une fois les protéines digérées en peptides, la phase de quantification peut débuter. Néanmoins, si la phase de digestion a permis de réduire la taille de protéines de sorte à ce qu'elles

soient analysables en spectrométrie de masse, de nombreux peptides ont désormais des masses très proches, ce qui peut mener à un brouillage du signal et empêcher la détection d'une partie d'entre eux. Pour cette raison, les peptides sont dans un premier temps séparés au cours d'une phase de chromatographie en phase liquide. Les peptides sont placés dans une solution avec un gradient croissant de solvant organique qui circule à un débit contrôlé. De cette manière, les peptides les plus hydrophiles migrent le plus rapidement, tandis qu'à l'inverse, ceux les plus hydrophobes migrent plus tardivement, lorsque la quantité de solvant organique devient plus importante.

Une fois les peptides séparés, ils arrivent à l'entrée du spectromètre de masse où ils sont tout d'abord ionisés. Il existe de nombreuses techniques pour réaliser la spectrométrie de masse. Ici nous n'évoquerons que les méthodes utilisées pour réaliser l'analyse des données protéomiques au cours de ma thèse, le Trapped Ion Mobility Spectrometry (TIMS) (Ridgeway 2018) et le Time Of Flight (TOF), utilisés par le timsTOF Pro MS (Bruker Daltonics, Germany).

Le TIMS permet d'ajouter une dimension d'analyse supplémentaire en plus de l'hydrophobicité et de la masse, en se basant sur le concept d'ion mobility. La mobilité d'un ion dépend de sa masse, charge et forme. Deux ions de même masse mais de formes différentes interagiront différemment avec d'autres particules : l'ion le plus replié interagira le moins, et aura donc une mobilité plus importante. Dans le TIMS, dans un premier temps un flux de gaz pousse les ions tandis qu'un champ électrique croissant les ralentit. Ces derniers se stabilisent au niveau où les deux forces sont égales. Ainsi, les ions avec la mobilité la plus petite sont les plus avancés dans le dispositif (ils sont plus facilement poussés par le flux de gaz). Dans un second temps le champ électrique est progressivement baissé, ce qui permet aux ions d'avancer séquentiellement dans le dispositif. La figure 13 résume ce fonctionnement.

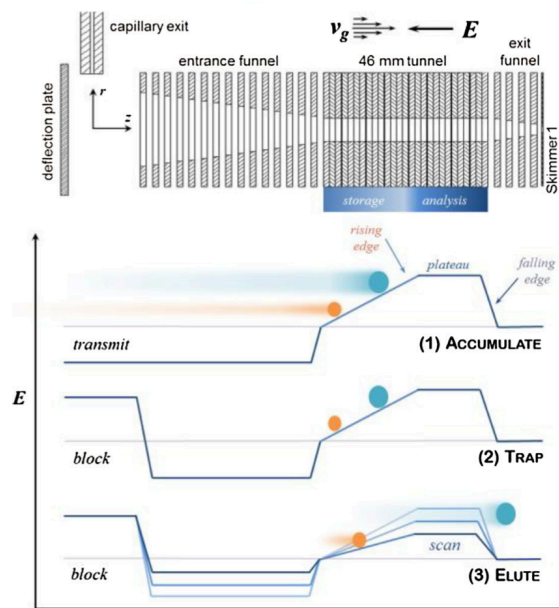


Figure 13 : Fonctionnement du TIMS. Haut : Les ions entrent dans un tunnel, poussés par un flux de gaz constant (V_g), et sont ralentis par un champ électrique E dont l'intensité augmente avec la distance parcourue. Bas : Dans un premier temps, dans l'état accumulation (accumulate), les ions entrent dans le tunnel et se stabilisent en fonction de leur mobilité. Dans un second temps, l'entrée d'autres ions est empêchée en imposant un fort champ électrique en entrée (état trap). Dans un troisième temps, l'intensité du champ électrique est progressivement diminuée, libérant les ions successivement suivant leur mobilité (état elute).

Ils sont ensuite accélérés avec des énergies égales pour traverser une tube sous vide, au bout duquel se situe un détecteur. Les peptides les plus légers sont les plus rapides, et sont donc détectés en premier. De leur temps de trajet est déduit leur ratio masse / charge (m/z), et du nombre de peptides détectés simultanément est déduite l'intensité (analyse TOF).

Acquisition des données

Les données brutes délivrées par le spectromètre de masse sont analysées informatiquement afin d'identifier les protéines présentes et de les quantifier.

Un logiciel spécialisé réalise informatiquement une digestion des protéines issues d'une base de données sélectionnée au préalable, en se rapprochant le plus possible des conditions de l'expérience, notamment en sélectionnant l'enzyme utilisée lors de la digestion des échantillons.

Les résultats obtenus informatiquement sont ensuite comparés à ceux issus de l'expérience afin d'identifier les peptides, puis à partir de ces derniers, les protéines.

De nombreux logiciels ont été développés dans ce but. En 2017, Verheggen et al (Verheggen 2017) ont publié une liste détaillant les outils disponibles, et en dénombraient près de 40.

Tous les logiciels ont besoin des informations suivantes pour fonctionner : une base de données de protéines à partir de laquelle réaliser la digestion, le nom de l'enzyme utilisée pour la digestion, le nombre maximum de lieux de clivages qui peuvent être manqués par l'enzyme lors de la digestion d'une protéine, les modifications d'acides aminés attendues du fait du pré-traitement des protéines, le niveau de charge attendu sur les peptides. Ces informations sont nécessaires afin de réaliser une digestion informatique la plus proche possible de celle qui a réellement lieu expérimentalement.

Afin de limiter au maximum les erreurs d'association entre les protéines informatiques et expérimentales, une approche par leurre a été développée (Elias 2007, Wang 2009, Navarro 2009). Au sein de la base de données de protéines, des leurres sont générés en créant des séquences d'acides aminés qui n'existent pas réellement. Puis les recherches d'identification sont réalisées par le logiciel en testant aussi les protéines leurres. Cela permet d'associer à chaque protéine expérimentale un taux de fausse découverte (FDR, pour False Discovery Rate), correspondant au taux de séquences associées à cette dernière provenant de leurres. Une FDR de moins de 1% est généralement prise comme critère dans les revues scientifiques.

4.1.2 PROTÉOMIQUE DANS LA MUCOVISCIDOSE

Les premières études protéomiques dans la mucoviscidose se sont concentrées sur CFTR, et plus particulièrement sur la recherche des protéines interagissant directement avec cette dernière, dans l'espoir de mieux comprendre le fonctionnement de CFTR au niveau cellulaire dans le cadre d'un réseau protéique. Par ailleurs, de telles études peuvent aussi permettre d'en apprendre davantage sur le mécanisme de repliement de CFTR et de transport de la protéine jusqu'à la membrane. La première étude protéomique visant à définir les interactants des protéines CFTR a été réalisée en 2006 (Wang 2006) : Wang et al. ont étudié l'interactome de CFTR non muté sur des lignées cellulaires pulmonaires et intestinales (BHK, Calu-3, HT29 et T84). Cette étude a ainsi permis de mettre en évidence 167 protéines interagissant avec CFTR. L'interactome de CFTR muté pour la mutation F508del a aussi été analysé dans cette étude sur la lignée cellulaire BHK. Il s'agit de la première étude protéomique différentielle dans le cadre de la mucoviscidose, comparant les

interactomes de CFTR muté et non muté. Etant donné que la protéine F508del-CFTR présente une altération de son repliement et de son transport à la membrane plasmique, l'analyse différentielle des interactants de CFTR a permis de mettre en évidence les voies qui sont activées différemment lorsque qu'un CFTR correctement ou incorrectement replié est transporté. Wang et al. ont ainsi constaté que la protéine CFTR, qu'elle soit mutée ou non, a de fortes interactions avec un ensemble de protéines chaperonnes et cochaperones, des protéines dont la fonction est d'assister d'autres protéines dans leur stabilisation et leur maturation. En particulier, les cochaperones Hsp90 apparaissent comme très importantes pour le repliement de CFTR en jouant sur la stabilité du repliement de cette dernière dans le RE. L'étude a ainsi révélé que l'interaction de la cochaperone Hsp90 Aha1 est différente pour CFTR non muté et pour CFTR F508del. La désactivation d'Hsp90 Aha1 a de plus mené à une augmentation de l'activité de CFTR F508del à la membrane plasmique, suggérant que cette protéine joue un rôle dans la dégradation de CFTR muté, et qu'elle pourrait potentiellement être une cible intéressante dans le cadre de la mucoviscidose. Il est par ailleurs intéressant de noter que l'incubation des cellules à 30°C permet d'augmenter la quantité de CFTR F508del qui atteint la membrane des cellules. A température réduite, la voie de repliement de F508del semble être plus stabilisée. La conclusion de cette étude est donc que l'activité des chaperonnes peut être ajustée pour augmenter ou pour diminuer le repliement de CFTR F508del avant son transport à la membrane plasmique. Ainsi, l'activité des chaperonnes peut être un mécanisme régulant le repliement pour la sortie du RE.

En 2015, une autre étude a identifié un ensemble d'interacteurs de CFTR F508del (Pankow 2015), en particulier ceux qui déclenchent potentiellement le phénotype de la maladie. L'analyse a été réalisée sur des lignées cellulaires épithéliales bronchiques immortalisées HBE14o-, exprimant CFTR non muté, et CFBE41o-, exprimant CFTR F508del. La comparaison des interactomes a mis en évidence un très haut niveau de similarité, avec plus de 85% d'identité (368 protéines interagissent à la fois avec CFTR muté et non muté). En outre, 208 et 62 protéines supplémentaires ont été mises en évidence comme interagissant exclusivement avec CFTR F508del et CFTR non muté respectivement. Parmi les interacteurs spécifiques de CFTR F508del, on retrouve les chaperones mises en évidence par Wang et al. (notamment Hsp90) ainsi que des protéines connues pour être impliquées dans la dégradation de CFTR, comme AMFR, STUB1 et VCP. De manière intéressante, l'étude a également mis en évidence un ensemble de protéines déjà connues pour être impliquées dans la voie de dégradation associée au RE des protéines mal repliées, mais qui n'avaient encore jamais été associées directement à la mucoviscidose. L'étude s'est aussi intéressée aux effets de différents mécanismes de sauvetage de CFTR F508del sur son interactome,

notamment l'incubation des cellules à une température basse (26-30°C). Le changement de température a considérablement modifié l'interactome de CFTR F508del (plus de 65% des interactants protéiques ont été modifiés). Les protéines modifiées peuvent être divisées en trois classes : les protéines impliquées dans la dégradation (appartenant notamment à la voie de dégradation associée au RE, et à l'élimination des protéines de la membrane plasmique) ; les protéines impliquées dans le traitement du repliement (telles que Hsp90 et les protéines régulées par le glucose) ; et les protéines impliquées dans le traitement de l'ARN (telles que PABPC1, impliquée dans le métabolisme de l'ARNm).

En 2018, une étude du groupe de Carlos Farinha (Farinha 2018) s'est intéressée plus particulièrement à la régulation de la rétention de F508del-CFTR dans le RE, contrôlée par le système de contrôle de qualité du RE. Les auteurs ont ainsi identifié les protéines qui interagissent spécifiquement avec CFTR au niveau de motifs importants pour les points de contrôle du RE. En particulier, ils se sont intéressés au motif arginine-framed tripeptide, utilisé par la cellule comme un signal négatif pour retenir les protéines non repliées dans le RE (Roxo-Rosa 2006), et au motif di-acide, nécessaire pour l'exportation de la protéine hors du RE (Farinha 2013). L'état de repliement de CFTR a un impact sur l'exposition de ces motifs, responsables soit de la rétention de CFTR dans le RE (et de sa dégradation ultérieure par le protéasome), soit de la sortie de CFTR du RE. Parmi les protéines interagissant fortement avec CFTR, les auteurs ont identifié des composants généraux des mécanismes de repliement et de protéostase, tels que des GEF pour le trafic, ou des GTPases comme CDC42BPG ou ARHGEF1, ce qui soutient l'idée que la reconnaissance de tels motifs est un mécanisme d'évaluation du repliement. Cette étude a ainsi été la première à introduire le concept selon lequel l'interactome de CFTR est bien plus façonné par l'état de repliement de la protéine (évalué par sa capacité à sortir du RE et à atteindre la membrane plasmique) que par sa localisation subcellulaire.

En 2019, Matos et al. (Matos 2019) ont réalisé la première étude de l'interactome de CFTR au niveau de la membrane plasmique. En utilisant une lignée de cellules bronchiques immortalisées modifiées (Botelho 2015), ils ont développé un protocole permettant de capturer sélectivement l'interactome de CFTR à la membrane plasmique. Ils ont ainsi étudié les interactomes de CFTR non muté et de CFTR F508del après traitement pharmacologique afin de trouver des différences dans les interactants de CFTR au niveau de la membrane plasmique. Leur étude a ainsi mis en évidence 22 protéines impliquées dans des complexes protéiques macromoléculaires localisés soit dans la membrane plasmique, soit dans le cytoplasme. Parmi elles, la calpaïne-1, une protéine appartenant aux complexes NHERF1-ezrine est apparue particulièrement intéressante du fait de son implication

connue dans la mucoviscidose. En effet, la calpaïne-1 est une protéine dont l'inhibition a favorisé le sauvetage du F508del-CFTR dans les cellules du sang périphérique de patients atteints par la mucoviscidose (Averna 2011). Les résultats de Matos et al. ont révélé le rôle de la calpaïne-1 en tant qu'interactant exclusif de CFTR F508del qui réduit la stabilité de ce dernier dans la membrane plasmique. Plus important encore, leur étude a également confirmé que la régulation négative de la calpaïne 1 ou son inhibition chimique améliorent considérablement le sauvetage fonctionnel du CFTR F508del dans les cellules des voies respiratoires.

La protéomique est également largement utilisée pour étudier les modifications post-traductionnelles. Dans le cadre de la mucoviscidose, l'étude de ces modifications dans les protéines mutantes par rapport aux protéines non mutées aide à comprendre le fonctionnement des mutants, et pourrait permettre le développement de nouvelles stratégies thérapeutiques. En 2019, Pankow et al. ont étudié comment certaines modifications post-traductionnelles (Pankow 2019) telles que la phosphorylation et la méthylation régulent le renouvellement de CFTR. Pour cela, ils ont vérifié si des modifications post-traductionnelles différentes étaient causées par les mutations de CFTR, et si ces différences étaient impliquées dans le sauvetage de CFTR après traitement. Il s'agit de la seule étude dans laquelle les modifications post-traductionnelles ont été mises en avant comme étant des régulateurs clés du trafic et de la biogénèse de CFTR. Dans un premier temps, les auteurs ont analysé ces modifications dans des lignées cellulaires 16HBE, exprimant CFTR non muté, et CFBE41o-, exprimant CFTR F508del, et sont ainsi parvenus à identifier 37 sites de modifications post-traductionnelles (20 phosphorylations, 6 méthylations et 11 ubiquitinations). Certaines de ces modifications post-traductionnelles n'avaient jamais été rapportées auparavant, comme les sites de méthylation, prédominants dans le cas de CFTR non muté et les sites d'ubiquitination, principalement détectés dans le cas de CFTR F508del. Ils ont ensuite quantifié certaines des modifications post-traductionnelles ainsi détectées, révélant une réduction de la méthylation de trois acides aminés (Lys442, Lys584 et Arg751) et une réduction de la phosphorylation de trois autres (Thr421, Ser422 et Ser427), dans le cas de CFTR F508del par rapport à CFTR non muté. En outre, les auteurs ont quantifié l'abondance de la phosphorylation chez quatre mutants de CFTR : G551D, R117H, N1303K et F508del. Leurs études ont montré que la quantité de phosphorylation dans les mutants G551D et R117H (ayant une activité de canal ionique défectueuse sans effets sur la biogénèse) était comparable à la quantité de phosphorylation dans le cas de CFTR non muté. Au contraire, la phosphorylation de Thr421 et Ser427 était drastiquement réduite pour les mutations N1303K et F508del (toutes deux associées à une maturation défectueuse), suggérant que ces sites de phosphorylation sont cruciaux pour la biogénèse de CFTR.

Cette étude est la première à souligner le rôle essentiel des modifications post-traductionnelles dans le repliement des protéines et le contrôle de la qualité au cours de la maturation de CFTR.

Les études protéomiques peuvent aussi être réalisées au niveau général de la cellule, en réalisant un inventaire des protéines présentes et de leurs quantités respectives. Actuellement, la technologie permet la quantification fiable de plusieurs milliers de protéines en une seule analyse (O'Connell 2018). Dans de telles études, les analyses comparatives sont réalisées en comparant les niveaux d'expression des protéines détectées dans les différentes conditions considérées. L'idée derrière ces analyses est que les protéines sont les effecteurs ultimes de la plupart des processus cellulaires, et les protéines exprimées différemment pourraient donc révéler la physiopathologie de la maladie. Très peu d'études ont été menées pour établir un profil d'expression des protéines dans les tissus respiratoires dans la mucoviscidose. En 2006, Frischer et al. ont réalisé un profil d'expression protéique global en analysant les biopsies bronchiques de neuf patients atteints par la mucoviscidose, et les ont comparés avec ceux de huit sujets témoins (Frischer 2006). Plus de 300 protéines ont ainsi été identifiées comme étant différentiellement exprimées dans les tissus bronchiques des individus atteints par la mucoviscidose par rapports aux contrôles sains. Parmi elles, des protéines impliquées dans l'inflammation, dans l'infection, et dans la réponse au stress cellulaire, ont été détectées. En 2018, en utilisant une approche similaire sur les lignées cellulaires CFBE41o- et 16HBE14, Puglia et al. ont identifié 15 protéines significativement dérégulées dans le cas de la présence de la mutation F508del (Puglia 2018). Cependant ces études utilisent des méthodes d'analyse ne détectant que les protéines les plus fortement exprimées dans la cellule.

En 2014, le groupe de Rauniyar a été le premier à essayer d'augmenter le nombre de protéines détectées lors d'une analyse (Rauniyar 2014). Ils ont ainsi analysé le protéome de CFBE41o- (cellules contenant CFTR F508del) en comparaison avec celui de lignées cellulaires 16HBE14o-. Le changement de technique d'analyse leur a permis d'étudier en profondeur le protéome, obtenant des informations sur des milliers de protéines possiblement importantes, soit pour comprendre le mécanisme moléculaire de la maladie, soit comme nouveaux marqueurs moléculaires. La groupe a ainsi identifié 349 protéines exprimées de manière différentielle (218 régulées à la hausse dans le cas de CFTR muté et 131 à la baisse), dont certaines appartiennent à des processus biologiques directement liés à la mucoviscidose, tels que la réponse aux protéines non repliées, la régulation de la protéolyse, le repliement des protéines et l'endocytose. Cependant, la limite principale de cette étude a été l'utilisation de lignées cellulaires, plus éloignées de la réalité

du fonctionnement cellulaire chez un patients que ne le sont des cellules primaires provenant des patients directement.

Ainsi, une étape supplémentaire dans la compréhension de l'impact des mutations de CFTR sur le protéome des cellules bronchiques a été rendue possible par l'utilisation de cellules épithéliales bronchiques humaines primaires qui reflètent directement la physiopathologie de la mucoviscidose, et notamment l'hétérogénéité des patients. Braccia et al. (Braccia 2019) ont pu identifier 154 protéines dérégulées dans la mucoviscidose (94 régulées à la hausse dans le cas de la mucoviscidose et 60 régulées à la baisse), y compris des interactants connus de CFTR, mais aussi des protéines dont le lien avec la mucoviscidose n'était pas connu auparavant.

Les différentes études protéomiques réalisées dans le cadre de la mucoviscidose ont montré qu'elles permettaient de mieux comprendre le fonctionnement de CFTR et son mécanisme de maturation, ainsi que de mettre en évidence de nouvelles cibles d'intérêt pour améliorer le niveau d'activation de CFTR muté. Cependant, à notre connaissance, aucune étude ne s'est intéressé aux effets des modulateurs de CFTR sur le profil protéomique de cellules respiratoires de patients atteints par la mucoviscidose. De plus, la grande majorité des études s'est concentrée uniquement sur la mutation F508del, laissant de côté toutes les mutations de classes I. Enfin, les rares analyses protéomiques globales n'ont pas étudié les différences d'expression au niveau des voies de signalisation, ce qui pourrait pourtant permettre d'obtenir une vue plus globale du fonctionnement des cellules contenant CFTR muté.

La protéomique apparait donc comme un outil très intéressant, dont le plein potentiel n'a pas encore été exploité dans la mucoviscidose. Son utilisation pour analyser des patients traités par modulateurs pourrait permettre de mieux comprendre le fonctionnement de ces derniers, et potentiellement de mettre en évidence de nouveaux bio marqueurs d'intérêt dans cette maladie.

L'étude 2 de ce manuscrit est consacrée à l'analyse protéomique de la sueur de patients atteints par la mucoviscidose. L'étude 3 contient quant à elle une analyse de cellules respiratoires de patients porteurs de mutations de classes I ou II, traités ou non par modulateurs, dans l'objectif de comprendre l'effet de ces derniers sur la cellule.

4.2 L'analyse transcriptomique

4.2.1 RÉALISATION PRATIQUE

La première étape d'une expérience RNA-seq est l'extraction et la purification de l'ARN. Les ARN sont d'abord séparés des ADN et des protéines. L'ARN total ainsi extrait contient aussi des ARN ribosomiques et des ARN de transfert qui ne sont généralement pas pertinents pour la recherche de l'expression des gènes. Par conséquent, dans le cas où la question biologique ne concerne que les ARNm, une étape de purification est réalisée en tirant avantage de la réaction de polyadénylation qui n'a lieu que pour ces derniers. Les ARNm viennent se fixer sur des queues poly(A) et sont ainsi retenus.

Après extraction et purification, les ARN sont fragmentés, et ces derniers sont ensuite soumis à une transcription inverse en ADN complémentaire (ADNc) double brin. Enfin, les ADNc ainsi produits sont amplifiés par PCR pour renforcer le signal final que l'on obtiendra.

Pour le séquençage Illumina, qui est la méthode la plus couramment utilisée aujourd'hui en transcriptomique mais aussi la méthode utilisée pour les études qui seront décrites dans ce manuscrit, les ADNc amplifiés sont liés à de courts oligonucléotides complémentaires des séquences adaptatrices elles mêmes fixées à une plaque, puis séquencés avec du désoxynucléoside triphosphate (dNTP) marqué par fluorescence, de manière progressive. Après l'ajout d'un dNTP, le marqueur fluorescent agit comme un terminateur et empêche ainsi l'ajout d'autres dNTP. Ensuite, une image est prise pour capturer les signaux de fluorescence et en déduire quels dNTP ont été ajoutés. Le marqueur fluorescent est ensuite clivé pour ajouter une autre couche de dNTP.

Cette étape de séquençage introduit de nombreuses erreurs. La terminaison par les marqueurs fluorescents n'est pas parfaite. Il n'est pas rare que plus d'un dNTP soit ajouté au cours d'une seule étape. L'appariement lui-même n'est pas parfait non plus. Ainsi, chaque image fluorescente contient du bruit. Cependant, comme les ADNc sont amplifiés avant le séquençage, cette erreur est largement réduite puisque chaque dNTP est déduit à partir de la moyenne des signaux obtenus. De cette façon également, un score de séquençage peut être évalué et enregistré pour chaque base (en regardant si le NTP varie d'un ADNc à un autre), ce qui permet de réaliser un filtrage informatique en aval.

Le séquençage Illumina peut être effectué selon deux modes différents : le séquençage à extrémité unique ou le séquençage à extrémité appariée, indiquant la ou les extrémités à partir desquels l'ARNc est séquencé.

La technique RNA-seq peut être appliquée à des échantillons de tissus comprenant de nombreuses cellules, ou à des cellules individuelles. Dans le premier cas, la RNA-seq séquence un mélange de cellules de chaque échantillon, tandis que la RNA-seq unicellulaire récemment développée applique des techniques de séparation cellulaire pour capturer et séquencer l'ARN dans des cellules individuelles (Stark 2019). Cependant, si les coûts d'utilisation de la première méthode ont diminué de manière drastique au cours de la dernière décennie, ce n'est pas le cas de l'analyse ayant recours à la seconde méthode. Pour cette raison, nous nous sommes concentrés sur des analyses RNA-seq multicellulaires au cours de cette thèse.

La conception expérimentale est essentielle avant toute expérience RNA-seq. Un point important est que si les échantillons sont traités en plusieurs lots, l'expérience doit limiter au maximum les disparités systématiques entre lots. Cela signifie notamment que chaque lot doit contenir toutes les conditions expérimentales. Cette précaution permet par la suite de fournir les informations nécessaires aux méthodes de calcul en aval pour réduire ces différences artéfactuelles entre les lots.

4.2.2 TRANSCRIPTOMIQUE DANS LA MUCOVISCIDOSE

De manière similaire à la protéomique, la recherche dans la mucoviscidose se focalisant essentiellement sur CFTR et ses interactions directes, peu d'études transcriptomiques ont été réalisées afin d'étudier le mécanisme global des cellules.

Le premier profilage transcriptomique de cultures primaires différenciées d'épithéliums des voies respiratoires humaines provenant de donneurs homozygotes pour la mutation F508del a été réalisé par Zabner et al. en 2005 (Zabner 2005) en utilisant la méthode des puces à ADN pour la sélection et la rétention des ARNs à analyser. Dans cette méthode, la première étape consiste à piéger les ARNs sur une plaque. Sur les 22 000 gènes représentés sur la puce utilisée dans l'étude, 18 ont été observés comme étant significativement sur exprimés dans la mucoviscidose, tandis que 6 étaient sous exprimés. Le cotransporteur KCl KCC4 a été identifié comme étant élevé dans la mucoviscidose, et a donc été considéré en tant que candidat pour des études supplémentaires. Il est intéressant de noter que ce profilage a conclu que le niveau de l'ARNm de CFTR n'était pas significativement différent dans les cellules homozygotes pour la mutation F508del par rapport aux cellules de patients sains. De plus, aucune différence notable n'a été observée en comparant les patients masculins et ceux féminins.

Par la suite, l'une des premières études transcriptomiques utilisant l'épithélium nasal de sujets atteints par la mucoviscidose a été réalisée par Wright et al. (Wright 2006). Dans cette dernière, les auteurs ont comparé des patients homozygotes pour la mutation F508del et fortement touchés au niveau pulmonaire à d'autres porteurs des mêmes mutations mais ne présentant des symptômes pulmonaires que très légers (mesuré par le VEMS). En parallèle, les patients homozygotes pour la mutation F508del ont également été comparés à des contrôles sains. Une différence d'expression significative a été observée chez les patients présentant une maladie pulmonaire sévère pour 569 gènes. En particulier, les gènes impliqués dans l'ubiquitination des protéines, l'activité oxydoréductase mitochondriale, et le métabolisme des lipides, étaient significativement enrichis. Parmi les gènes sous exprimés chez les patients sévères, on trouve DUOX2, un producteur clé de peroxyde d'hydrogène pour la défense de la muqueuse des voies respiratoires, et la calréticuline, une chaperonne du RE impliquée dans le métabolisme des protéines.

La réaction en chaîne par polymérase avec rétrotranscription (RT-PCR) réalisée dans cette étude n'a cependant pas révélé de différences significatives dans les niveaux d'expression transcriptomiques de CFTR entre les patients homozygotes pour la mutation F508del et les contrôles sains, ce qui est en accord avec les résultats de Zabner et al.

En 2013, Clarke et ses collègues ont réalisé une étude par puces à ADN du génome entier de cellules épithéliales nasales primaires provenant d'homozygotes F508del et de contrôles sains, et ont comparé leurs résultats avec les autres études transcriptomiques déjà menées dans la mucoviscidose (Clarke 2013). Dans leur profil d'expression, les gènes impliqués dans la prolifération cellulaire étaient significativement sur-exprimés dans la mucoviscidose, tandis que les gènes liés aux cils cellulaires étaient sous-exprimés. En raison de la grande variabilité des profils d'expression génique entre les études, la méta-analyse comparant cette étude à cinq autres expériences de puces à ARN (y compris les études de Zabner et de Wright décrites précédemment) n'a mis en avant que très peu de gènes dérégulés communs à au moins trois expériences. Toutefois, lorsque les auteurs ont comparé les résultats de leurs puces à ADN à ceux de la puce d'Ogilvie (Ogilvie 2011), l'étude la plus récente au moment des analyses de Clarke, ils ont observé 21 gènes communs sur-exprimés et 9 gènes communs sous-exprimés. De manière intéressante, un nombre important de ces gènes étaient impliqués dans l'inflammation et la défense, notamment les transcrits CXCR4, FOS, S100A8, S100A9 et SERPINA3, tous régulés à la hausse.

Polineni et al. ont réalisé la plus grande analyse transcriptomique menée sur des patients atteints par la mucoviscidose. Il ont ainsi séquencé l'ARN des cellules de la muqueuse nasale de 134

sujets atteints par la mucoviscidose, dont le génotype et la gravité de la maladie variaient (Polineni 2018). L'analyse des voies d'expression des gènes a mis en évidence la corrélation positive entre la gravité de la maladie et l'infection virale, la signalisation inflammatoire, le métabolisme des lipides, la fonction des macrophages et l'immunité innée. Plusieurs gènes du complexe majeur d'histocompatibilité ont fortement contribué aux voies enrichies. Les auteurs ont conclu que les gènes de ce complexe pourraient servir de cibles dans le cadre de la mucoviscidose.

Une autre étude transcriptomique comparant les leucocytes du sang périphérique de sujets atteints par la mucoviscidose et présentant une maladie pulmonaire légère ou sévère a été réalisée par Kormann et al (Kormann 2017). L'analyse a permis d'identifier les gènes de la réponse à l'interféron de type I, ainsi que les protéines de la tige ribosomique, comme étant sur-exprimés chez les patients avec une maladie pulmonaire légère. Ces gènes apparaissent donc comme de potentiels modificateurs de la maladie pulmonaire dans la mucoviscidose et pourraient avoir des implications en tant que potentiels nouveaux biomarqueurs ou en tant que nouvelles cibles.

La baisse des coûts de réalisation d'analyses transcriptomiques a permis une augmentation du nombre d'études réalisées dans ce domaine ces dernières années. En 2020, Ling et ses collègues (Ling 2020) ont ainsi étudié le transcriptome de cellules primaires épithéliales provenant de 7 patients homozygotes pour la mutation F508del et de 5 contrôles sains, en les infectant ou non avec le rhinovirus. Ils ont ainsi pu constater que les malades de la mucoviscidose et les contrôles sains présentaient des réponses similaires à l'infection, mais que la réponse à l'infection était plus complexe chez les cellules provenant des patients que chez les contrôles sains, avec plusieurs voies biologiques surreprésentées, telles que la signalisation des cytokines et les voies du métabolisme.

La même année, Saint-Criq et son équipe (Saint-Criq 2020) se sont intéressés à l'influence du milieu de culture sur le profil transcriptomique des cellules de patients atteints par la mucoviscidose, et notamment sur son importance sur la réponse aux modulateurs. Ils concluent que le milieu de culture joue un rôle prépondérant sur la différenciation et sur le profil d'expression transcriptomique, mettant ainsi en avant l'importance d'avoir un milieu de culture homogène lors de la réalisation d'une étude transcriptomique pour la mucoviscidose.

Enfin, en 2021, Rehman et al. (Rehman 2021) ont étudié l'effet de l'ajout de marqueurs pro inflammatoires sur le transcriptome de cellules de patients atteints par la mucoviscidose. Les données analysées ont mis en lumière une corrélation positive entre l'inflammation des voies respiratoires et l'amélioration des fonctions pulmonaires dues aux modulateurs de CFTR : l'effet des modulateurs était en effet plus important dans les cellules exposées à des marqueurs pro inflammatoires.

Les différentes études évoquées dans cette section témoignent de l'utilité des analyses transcriptomiques dans le cadre de la mucoviscidose. Elles permettent non seulement de mettre en lumière le fonctionnement et les interactions de CFTR muté ou non au sein des cellules épithéliales en réponse à différents stimuli d'intérêt (signal pro inflammatoire, infection...), mais elles identifient aussi de nouvelles cibles potentielles d'intérêt dont l'activation ou l'inhibition pourrait mener à une amélioration des conditions de vie des patients atteints par la mucoviscidose. Pourtant, à notre connaissance, aucune étude ne s'est encore intéressée à l'effet des modulateurs sur le profil transcriptomique des cellules épithéliales de patients atteints par la mucoviscidose (seule l'équipe de Saint-Criq s'y est intéressée à la marge dans son étude sur les milieux de culture). De telles études permettraient de mieux comprendre le mécanisme d'action de ces derniers, et potentiellement de mettre en évidence des gènes/protéines différents de CFTR ciblés par les modulateurs. Dans cet objectif, j'ai réalisé au cours de ma thèse une analyse transcriptomique de cellules provenant de patients porteurs de mutations de classe II, traitées ou non avec des modulateurs de CFTR (étude 3 de ce manuscrit).

Un autre point important soulevé par ces études est la grande hétérogénéité des résultats obtenus lorsque l'analyse porte sur l'expression des gènes, malgré des conditions de départ similaires. En revanche, ces mêmes résultats apparaissent beaucoup plus proches à travers les différentes études lorsqu'ils sont analysés au niveau des voies de signalisation (au niveau de groupes de gènes). Néanmoins, au delà de la méthode GSEA (pour Gene Set Enrichment Analysis), utilisée par la plupart des études évoquées précédemment, peu de techniques existent pour réaliser une telle analyse par voies. Au cours de cette thèse, je me suis donc intéressé au développement d'un nouvel outil dédié à l'analyse de groupements de gènes dans des analyses omiques, basé sur l'analyse en composantes principales (ACP), appelé rRoma. Afin de comprendre l'intérêt et le fonctionnement de cet algorithme, qui sera développé dans la section résultats de ce manuscrit, il est dans un premier temps nécessaire de comprendre le concept d'ACP. Celui-ci est expliqué dans la section méthodes et algorithmes suivante, aux côtés d'autres algorithmes et concepts d'intérêt que j'ai été amenés à utiliser au cours de ma thèse.

5. Méthodes et algorithmes employés

5.1 L'Analyse en Composantes Principales

L'analyse de données, quelle que soit la nature de ces dernières, demande bien souvent un processus de simplification de leur représentation afin d'en extraire des informations interprétables. Je me suis retrouvé face à cette nécessité par deux fois au cours de ma thèse, pour deux problématique très différentes.

Lors de mon analyse de l'évolution des paramètres cliniques en réponse au traitement par modulateurs (étude 1), il était nécessairement de pouvoir visualiser facilement et identifier les interdépendances des 6 paramètres pris en compte. En parallèle, le coeur de la problématique lors du développement de l'algorithme rRoma (étude 4) consistait à simplifier au maximum l'information complexe contenue dans les niveaux d'expression de différents gènes aux fonctions similaires ou appartenant aux mêmes voies de signalisation.

La base mathématique pour résoudre ces deux problématiques a été la même. J'ai utilisé des variations de la méthode d'ACP (Pearson 1901, Hotelling 1933). Il s'agit d'une méthode permettant de réduire la dimension des données (donc, le nombre de paramètres considérés) en conservant le maximum d'information possible, et de mettre en évidence les relations entre les différents paramètres pris en compte dans l'étude.

Dans cette section, j'explique en détail le fonctionnement de la méthode ACP, dans un premier temps au moyen d'une approche simplifiée pour les non mathématiciens, et dans un second temps d'un point de vue mathématique. Ces explications permettront de mieux comprendre les analyses conduites dans les études 1 et 4 de ce manuscrit.

5.1.1 EXPLICATION POUR LES NON MATHÉMATIENS

Cas simple de deux dimensions

Dans un premier temps, considérons une cohorte de patients qui serait décrite par deux paramètres cliniques. La cohorte est notre jeu de données, et comme il n'y a que deux paramètres, les patients peuvent être représentés par des points dans un espace de deux dimensions, c'est-à-dire

un plan (l'horizontale représentant le premier paramètres clinique, et la verticale le second) (figure 14, panel de gauche)

Dans une ACP, deux combinaisons des 2 paramètres initiaux sont employées, appelées composantes principales 1 et 2 (PC1 et PC2), et les données sont représentées selon ces nouveaux paramètres correspondant à de nouveaux axes horizontaux et verticaux. Les composantes PC1 et PC2 sont calculées de telle sorte que les données soient le plus dispersées possibles le long de l'axe PC1, la variance entre les points étant beaucoup plus faible selon l'axe PC2.

On peut alors utiliser cette nouvelle représentation pour réduire la dimension de nos données. Si l'on décidait de représenter des points de données le long d'une droite unique tout en conservant le plus de variance possible, alors la meilleure droite correspondrait à l'axe de la PC1. On perdrait certes l'information (de la variance) liée à la PC2, mais cette dernière est par construction beaucoup plus faible.

Cet exemple reste limité : passer de 2 dimension à une seule n'est pas d'une grande utilité puisque l'on peut très bien visualiser les informations en 2 dimensions. Mais il permet de poser les concepts pour les cas plus complexes pour lesquels plus de paramètres sont pris en compte, et donc pour lesquels le nombre de dimensions de départ est plus grand.

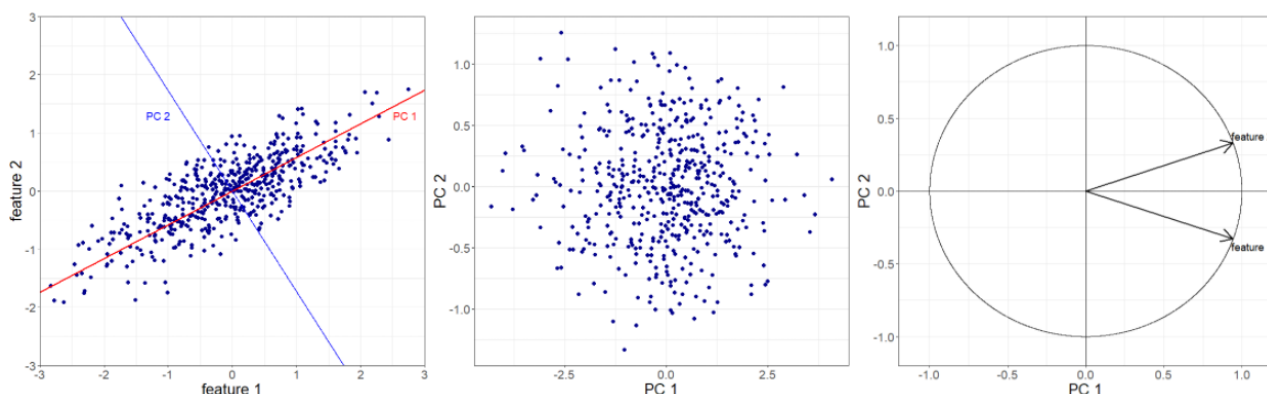


Figure 14 : Visualisation de l'effet de l'ACP dans un cas à deux dimensions. Gauche : nuage de points initial. La PC1 est y est représentée en rouge et capture le maximum de variance de la distribution. La PC2 est représentée en bleu. Milieu : Représentation du nuage de points dans l'espace des deux premières composantes. Droite : Cercle de participation des dimensions initiales aux composantes principales

Cas de 3 dimensions

Pour mieux comprendre l'importance de l'ACP, plaçons-nous désormais dans le cas à trois dimensions, le plus simple pour lequel la réduction de dimension est utile. Reprenons l'exemple précédent, en ajoutant un troisième paramètre pour décrire les patients. Ces derniers doivent donc désormais être représentés par des points projetés dans un espace à trois dimensions, plus difficilement lisible (figure 15 gauche).

Dans ce cas, l'ACP fonctionne de la même manière qu'en deux dimensions. On oriente l'espace (on le « fait tourner ») en définissant trois nouvelles directions perpendiculaires les unes aux autres (les PC1, PC2 et PC3), de sorte à ce que la variance le long de la première direction (la PC1) soit la plus grande possible, puis que la variance restante soit la plus grande possible le long de la PC2, la PC3 capturant la variance restante non prise en compte dans les PC1 et PC2.

Dans cette nouvelle représentation, la plus grande partie de la dispersion des points est visible le long des PC1 et PC2. Pour simplifier la représentation, on peut alors représenter les données en 2 dimensions, i.e. dans le plan défini par PC1 et PC2, en minimisant l'information que l'on ne prend pas en compte, c'est-à-dire celle portée par PC3 (figure 15 droite). Le pourcentage de la variance totale (donc de l'information totale) conservé par chaque PC est aussi indiqué. Ainsi, dans notre exemple, on a réalisé une réduction de dimension, de 3 à 2, en ne perdant que 15% de l'information présente initialement dans nos données.

Mais l'ACP permet aussi de déterminer quels paramètres de départ sont les plus importants pour expliquer la variance présente dans les données, et si les patients se répartissent de manière similaire selon certains paramètres (c'est-à-dire si les paramètres sont corrélés entre eux). Les PCs sont en fait obtenues en réalisant une combinaison linéaire des paramètres initiaux utilisés pour décrire les données. Un paramètre jouant un rôle important dans une PC aura un coefficient élevé, à l'inverse il sera très faible si le paramètre ne contribue que faiblement à cette composante. Ces coefficients témoignent donc aussi de la corrélation des paramètres avec les PCs : si plusieurs paramètres ont des coefficients élevés pour la même PC, cela signifie qu'il sont corrélés à cette PC, et donc entre eux.

L'information sur la participation des paramètres aux différentes PCs peut être visualisée dans l'espace de ces dernières, par des flèches représentant les paramètres de départ : plus une flèche est alignée avec une PC, plus elle contribue de manière importante à cette dernière (son coefficient pour cette composante est grand). Ainsi, dans notre exemple, on peut voir que les paramètres x et y participent de manière similaire et importante à la PC1, alors que le paramètre z

participe à la PC2. Les paramètres x et y sont donc corrélés, ils portaient une information redondante synthétisée par la PC1.

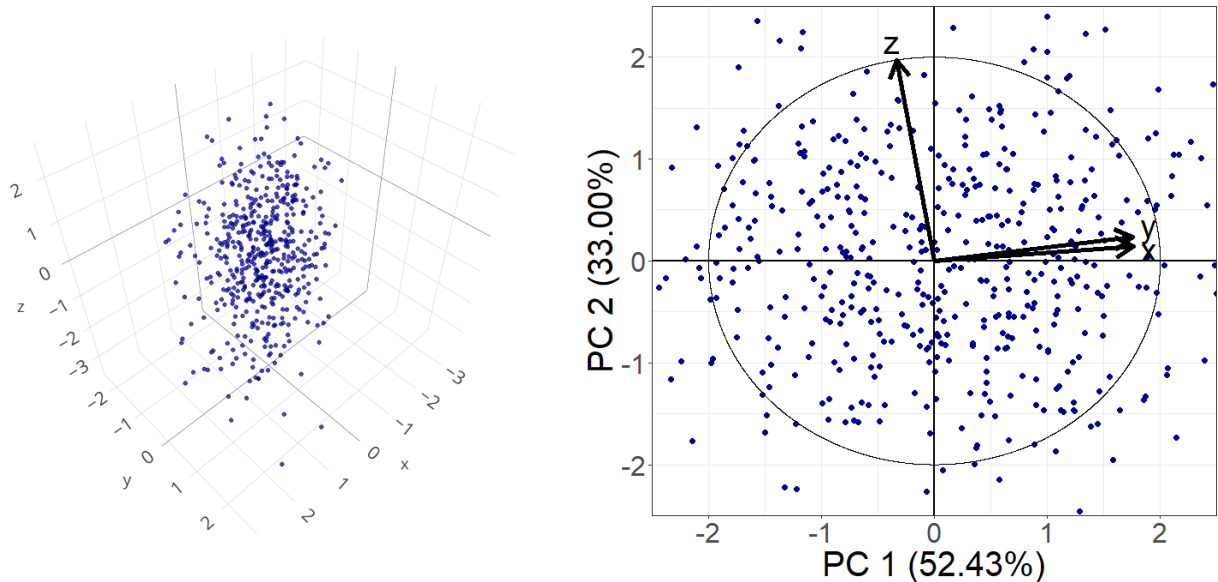


Figure 15 : Visualisation de l'ACP dans le cas d'une réduction de dimension de 3 à 2 composantes. Gauche : Représentation initiale du nuage de points. Droite : Nuage de points projeté dans l'espace des deux premières composantes principales, et cercle de participation des dimensions initiales à ces dernières.

Cas de plus de dimensions

L'ACP est généralement appliquée à des cas plus complexes pour lesquels les patients sont décrits par un plus grand nombre n de paramètres. On définit alors n nouvelles composantes, nommées PC1 à PC n , toutes perpendiculaires les unes aux autres. Elles sont organisées de sorte à ce que les points soient le plus dispersés possible le long de la PC1, puis le long de la PC2, et ainsi de suite jusqu'à la PC n , qui explique le moins de variance. Dans ces cas complexes, l'ACP permet donc de grandement simplifier l'information, par exemple en ne conservant que les deux premières PCs pour pouvoir facilement visualiser les données sur un plan. De manière plus générale, la réduction de dimension facilite l'interprétation des données et aide à en tirer des conclusions. De plus, comme pour le cas à trois dimensions, on peut déterminer quels paramètres initiaux sont le plus impliqués dans la construction des premières PCs et sont donc les plus intéressants pour l'étude. Enfin, on peut déterminer quelles corrélations existent entre les différents paramètres de départ : des paramètres participant fortement à la même PC sont corrélés, alors qu'à l'inverse deux paramètres qui ne contribuent jamais à la même PC ne le sont pas.

5.1.2 DÉTAILS MATHÉMATIQUES DE L'ACP

On définit un dataset constitué de n individus, décrits par p variables numériques, que l'on représente par une matrice X de taille $n \times p$. Dans cette matrice, la i ème colonne correspond au vecteur x_i des observations de la variable i . L'ACP doit déterminer quelle combinaison linéaire des x_i est associée au maximum de variance, c'est-à-dire trouver a tel que $var(Xa)$ soit maximal, avec $a = a_1, a_2, \dots, a_p$ et $Xa = \sum_{i=1}^p a_i x_i$.

Dans un tel cas, on a $var(Xa) = a^T C a$, où C est la matrice de variance covariance de X . Pour s'assurer que ce problème de maximisation ait une solution, il faut ajouter une condition sur a . La condition la plus simple est d'imposer que $\|a\| = 1$. Alors $a^T a = 1$.

On peut alors utiliser la méthode du multiplicateur de Lagrange pour trouver le maximum (Astaiza-Gómez 2020). On définit L le lagrangien tel que $L(x, \lambda) = x^T C x - \lambda(x^T x - 1)$. Si a est un maximum de $var(Xa)$ avec $\|a\| = 1$, alors il existe λ_0 tel que L admette une différentielle nulle en (a, λ_0) .

En particulier, pour tout $i \in \{1, \dots, p\}$:

$$\frac{\partial L}{\partial u_i} = 2 \frac{\partial a^T}{\partial u_i} C a = 2 \lambda_0 a_i$$

Or $\frac{\partial a^T}{\partial u_i} C$ correspond à la i ème ligne de C . Ces contraintes sont donc équivalentes à :

$$C a = \lambda_0 a$$

a est donc un vecteur propre de C , et λ_0 la valeur propre associée. De plus :

$$var(Xa) = a^T C a = \lambda_0 a^T a = \lambda_0 \|a\| = \lambda_0.$$

Le maximum est donc atteint lorsque λ_0 est la plus grande valeur propre de C . Ainsi, la projection associée au maximum de variance correspond au vecteur propre de la matrice de variance covariance associé à la plus grande valeur propre de cette dernière. C'est la première composante principale (Jolliffe 2016).

Les composantes suivantes sont obtenues de manières similaires en tant que vecteurs propres associés aux valeurs propres suivantes de la matrice de variance-covariance.

La trace de la matrice de variance covariance est égale à la somme des variances de tous les paramètres initiaux. Or elle est aussi égale à la somme de ses valeurs propres, donc à la somme des variances de la projection du nuage de point dans chaque composante principale. Grâce à cette

propriété, il est possible de savoir quelle proportion de la variance totale est expliquée par chaque composante. Si l'on note q_i la proportion de variance expliquée par la composante i , alors on a :

$$q_i = \frac{\lambda_i}{\sum_{j=1}^p \lambda_j}$$

Où $\{\lambda_1, \dots, \lambda_p\}$ sont les valeurs propres de la matrice de variance covariance de X , organisées en ordre décroissant.

5.2 L'algorithme de clustering k-means

Au cours de ma thèse, j'ai été amené à séparer des échantillons en plusieurs groupes avec des caractéristiques similaires. Pour cela, j'ai utilisé la méthode dite de k-means clustering.

Nos échantillons peuvent être représentés par des points dans un espace de dimension k (par exemple le nombre de caractéristiques prises en compte dans nos échantillons). On définit dans un premier temps un nombre fixe de centroïdes, des points placés aléatoirement dans cet espace. Un processus itératif débute alors :

- Chaque échantillon est associé au centroïde dont il est le plus proche
- La position des centroïdes est recalculée comme étant la position moyenne des échantillons qui lui sont reliés

Ces deux étapes se répètent jusqu'à convergence (i.e les centroïdes ne se déplacent quasiment plus). Les échantillons reliés au même centroïde sont alors considérés comme appartenant au même cluster.

5.3 Tests statistiques utilisés

Si les tests statistiques utilisés en biologie pour comparer des résultats entre deux groupes sont le plus souvent des tests T de Student (éventuellement appariés si les échantillons comparés proviennent des mêmes individus, mais sont prélevés à des instants différents), ces derniers présentent deux limitations majeures. D'une part, une hypothèse forte est la normalité de la distribution des données étudiées. Or il m'est arrivé de travailler avec des données ne respectant pas cette condition. D'autre part, lorsque de nombreux tests sont réalisés sur les mêmes échantillons, ils

est attendu que certains d'entre eux témoignent d'une différence significative en réalité inexistante. Afin de prendre en compte ce biais, il est nécessaire de réaliser une correction pour tests multiples.

5.3.1 CAS OÙ L'HYPOTHÈSE DE NORMALITÉ N'EST PAS VÉRIFIÉE

Dans les cas où l'hypothèse de normalité n'était pas vérifiée, le test qui a été employé est celui dit des rangs signés de Wilcoxon. Il s'agit d'une alternative non paramétrique au test de Student pour des échantillons appariés. L'hypothèse nulle est l'égalité des médianes entre les deux groupes.

On mesure pour chaque sujet de l'étude la différence entre la première et la deuxième mesure, puis les patients sont ordonnés par ordre croissant de la valeur absolue de cette différence, associant ainsi un rang à chaque patient. De même, à chaque patient est associé un signe, selon que la différence est négative ou positive. Les rangs sont ainsi « signés ». Ces rangs signés sont sommés, et l'hypothèse nulle est rejetée si la somme est trop éloignée de zéro.

Ce test se basant sur les rangs, il peut être employé sans vérification des conditions de normalité.

5.3.2 CAS DE TESTS MULTIPLES

On utilise généralement 0.05 comme limite à un test statistique pour déterminer si une différence est significative ou non. Cela signifie que l'on accepte un taux d'erreur de 5%. Ainsi, pour 100 tests statistiques différents réalisés à partir des mêmes échantillons et significatifs, en moyenne 5 sont en réalité des faux positifs. Des corrections pour tests multiples doivent ainsi être appliquées afin de limiter au maximum ces erreurs.

Ici sera décrite la correction de Benjamini-Hochberg, utilisée dans ce manuscrit, mais d'autres méthodes existent, notamment basées sur des permutations multiples du jeu de données initial.

La procédure de Benjamini-Hochberg consiste dans un premier temps à ordonner les différents tests réalisés par ordre croissant de p-values, et ainsi à leur associer un rang j . En notant n_{bp} le nombre de tests réalisés au total, les p-values de chaque test sont alors modifiées par la formule suivante : $p_{BenjaminiHochberg} = \min(p * n_{bp}/j, 1)$

On appellera en général q-value cette p-value ainsi modifiée, qui prend en compte la multiplicité des tests effectués, et on appliquera la même limite de 0.05 à partir de cette q-value pour définir si un test est effectivement significatif.

II. Objectifs de la thèse

Ce travail de thèse a eu pour objectif de mieux comprendre les effets des modulateurs de CFTR selon deux aspects : d'un côté, en s'intéressant à la réponse clinique des patients à ces molécules, et de l'autre, en étudiant l'effet de ces modulateurs à l'échelle cellulaire, en s'intéressant plus particulièrement à un biofluide, la sueur, et à des cellules épithéliales des voies respiratoires, particulièrement affectées dans la mucoviscidose.

Nous avons privilégié des approches innovantes pour ces deux volets du travail de thèse. Ainsi, au niveau clinique, nous avons travaillé plus particulièrement sur la caractérisation de l'hétérogénéité de la réponse aux traitements. En ce qui concerne les analyses omiques, nous avons réalisé la première étude protéomique de la sueur dans la mucoviscidose, et nous avons étudié les effets de la trithérapie sur des cellules respiratoires de patients porteurs de mutations de classe II, mais aussi de patients porteurs de mutations de classe I. Enfin, pour réaliser ces études, nous avons utilisé un nouvel outil informatique que nous avons développé, dédié à l'analyse de voies dans les études omiques, à partir d'une première version publiée en 2016 (Martignetti 2016).

Ce travail de thèse a donc traité des questions suivantes :

1. Peut-on caractériser l'hétérogénéité de la réponse clinique aux modulateurs de CFTR ?

Tous les patients ne répondent pas de la même manière à des traitements par modulateurs.

- Peut-on réduire la réponse clinique dans la mucoviscidose à l'évaluation d'un seul paramètre ?
- A l'inverse, peut-on mettre en évidence plusieurs profils de réponse à un traitement par modulateurs ? Quels seraient les paramètres d'intérêt dans ces groupes ?
- Les paramètres d'intérêt peuvent-ils varier avec l'âge des patients ?

La première étude décrite dans ce manuscrit tente de répondre à ces questions. La collecte des données provenant des patients a été réalisée par des collaborateurs, puis j'ai participé à l'intégralité des analyses réalisées dans cette étude.

2. Peut-on mieux comprendre le mécanisme d'action des modulateurs de CFTR à travers l'étude de la sueur ?

La sueur est un biofluide fondamental dans la mucoviscidose, à l'origine du diagnostic de la maladie. Nous avons réalisé une analyse protéomique poussée pour obtenir de nouvelles informations sur la maladie :

- Peut-on expliquer certains symptômes de la mucoviscidose grâce à l'analyse protéomique de la sueur ?

- Les modulateurs de CFTR ont-ils des effets visibles dans la sueur ?

- La sueur pourrait-elle contenir de nouveaux biomarqueurs caractéristiques de l'effet d'un traitement par modulateurs ?

Ces questions correspondent à la deuxième étude de ce manuscrit. Les données ont été collectées et analysées au spectromètre de masse par des collaborateurs, puis j'ai réalisé l'intégralité des analyses informatiques de cette étude.

3. Peut-on mieux comprendre le mécanisme d'action des modulateurs de CFTR à travers l'étude de cellules épithéliales respiratoire ?

La trithérapie VX-770/VX-661/VX-445 a drastiquement amélioré les conditions de vie des patients atteints par la mucoviscidose. Pourtant son mécanisme d'action dans les cellules respiratoires n'est pas totalement compris. Ses effets sur les patients porteurs de mutations de classes I, ne produisant pas CFTR, n'ont pas non plus été évalués :

- Quelles sont les mécanismes d'action de la trithérapie VX-770/VX-661/VX-445 dans les cellules épithéliales respiratoires ?

- Certains mécanismes de correction pourraient-ils être indépendants de CFTR ? Les modulateurs de CFTR pourraient-ils alors être utilisés chez les patients porteurs de mutations de classe I ?

Ces questions seront décrites dans la troisième partie de la section des résultats. Pour cette étude, j'ai participé à la culture des cellules et à l'extraction de l'ARN et des protéines. L'obtention des données brutes par RNA-seq et par spectrométrie de masse a été réalisée par des collaborateurs, puis j'ai réalisé l'intégralité des analyses informatiques à partir de ces dernières.

4. Quels outils utiliser lorsqu'une analyse gène par gène est insuffisante ?

Le nombre limité de patients dans les études protéomiques et transcriptomiques mène à un faible nombre de gènes d'intérêt détectés. Une approche permettant de contourner cette limitation est l'analyse de groupes de gènes. Cependant de nouveaux outils permettant de prendre en compte les comportements opposés au sein d'un même groupe de gène doivent être développés :

- Comment réaliser une analyse efficace par groupes de gènes ?
- Comment considérer des groupes de gènes comportant à la fois des activateurs et des inhibiteurs d'une même voie ?

Un package R a été développé dans l'optique de résoudre ces problématiques et sera décrit dans la partie quatre des résultats. Nous nous sommes basés sur un programme préexistant écrit en Java, et j'ai participé à l'intégralité des étapes d'amélioration de celui-ci.

III. Résultats

1. Analyse de la réponse clinique aux traitements par modulateurs de CFTR

L'article associé à cette section est publié dans *Pediatric Pulmonology* et donné en annexe de ce manuscrit.

1.1 Introduction

Etant donnée l'efficacité des modulateurs de CFTR, de plus en plus d'enfants vont être amenés à bénéficier de ces traitements. Comme évoqué en introduction de cette thèse, il est attendu que ces médicaments, s'ils sont pris par les patients dès leur plus jeune âge, pourraient au minimum ralentir l'évolution de la maladie chez ces derniers (Accurso 2018). Néanmoins, les effets de ces modulateurs de CFTR deviennent de plus en plus difficiles à évaluer à mesure que la recherche dans le domaine progresse. L'évaluation de ces effets s'avère particulièrement délicate chez les enfants, dont la plupart ne présentent que des symptômes légers de la mucoviscidose, voire sont asymptomatiques pour les plus jeunes d'entre eux (Stanojevic 2016). Pour tenter de pallier ce problème, de plus en plus de paramètres sont pris en considération, mais leur mesure requiert souvent des matériels qui ne sont pas disponibles dans tous les centres médicaux, rendant difficile une standardisation du suivi de ces paramètres. Citons notamment l'exemple du LCI, dont la mesure semble être indicative de symptômes chez les très jeunes patients atteints par la mucoviscidose, et qui s'améliorerait avec l'administration de modulateurs de CFTR efficaces, témoignant donc d'un bénéfice chez les jeunes patients (Ratjen 2018). Malheureusement, cette mesure requiert du matériel complexe, ainsi qu'une expertise associée, qui ne sont pas disponibles dans tous les centres médicaux.

En parallèle, les paramètres mesurés en routine en clinique montrent eux aussi aujourd'hui leurs limites. Si chez des patients adultes dont les fonctions respiratoires sont déjà détériorées, l'évolution des paramètres respiratoires, notamment le VEMS, le CVF et le DEM 25-75, permet de mettre en évidence un effet bénéfique des traitements par modulateurs pour ces patients (Sagel 2021, Burgel 2020, Corey 1997), l'utilité de ces mesures apparaît beaucoup plus limitée chez les enfants, qui présentent le plus souvent des valeurs de paramètres respiratoires normales pour leur âge. Le même constat peut être dressé pour la mesure de l'IMC, ou encore pour la quantité

d'antibiotiques administrés. Enfin, plusieurs études ont montré que la mesure de l'évolution de la concentration en ions chlorure dans la sueur des patients n'était pas corrélée à une amélioration clinique (Barry 2014, Durmowicz 2019, Masson 2019).

La question de la mise en place d'un score simple, évaluable facilement à partir de mesures standardisées et réalisables dans tout centre clinique, dont l'évolution serait associée à un effet bénéfique des modulateurs de CFTR pour le patient, et ce, quel que soit l'âge de ce dernier, n'a donc toujours pas été résolue.

Dans cette thèse nous avons privilégié une approche nouvelle. Jusqu'à présent les études se concentraient sur la recherche d'un score global, qui permettrait d'évaluer une réponse au niveau global pour une cohorte donnée. En particulier, la plupart des études cliniques évaluent les effets d'un traitement à partir de l'évolution moyenne d'un paramètre donné dans la cohorte, sans prendre en compte la potentielle hétérogénéité de cette dernière. A l'inverse, nous avons recherché s'il était possible de déterminer des sous-cohortes de patients qui évolueraient de manière similaire selon certains paramètres cliniques. Dans notre étude, partant du principe que les paramètres décrits précédemment ne devraient pas être considérés de manière indépendante puisqu'ils reflètent tous des conséquences du dysfonctionnement de CFTR dans différents tissus, nous souhaitons notamment mettre en lumière plusieurs profils de réponse, associés à des combinaisons différentes de ces paramètres cliniques. Une seule autre étude s'est intéressée de manière similaire à l'hétérogénéité de la réponse à des traitements dans le cas de broncheectasies (Sibila 2021), en étudiant les évolutions des valeurs de VEMS, du nombre d'exacerbations et du score de qualité de vie évalué par les patients. Cette dernière a montré une absence de corrélation entre l'évolution de ces trois paramètres, et surtout une très grande hétérogénéité dans l'évolution de ces paramètres au sein de la cohorte étudiée, suggérant l'existence de plusieurs sous-groupes réagissant différemment au traitement, et supportant donc notre approche.

En plus de déterminer des sous cohortes de patients, nous avons utilisé une approche mathématique appelée ACP, décrite en introduction de cette thèse et à nouveau évoquée brièvement ci-après, permettant de mettre en évidence d'éventuelles redondances dans l'évolution des différents paramètres évalués, et donc de ne conserver que ceux les plus importants pour la mise en évidence d'une réponse clinique aux modulateurs. De plus, l'ACP permet de déterminer quelle combinaison de paramètres explique la plus de variabilité dans les données, et donc, dans notre cas, de caractériser la variabilité des types de réponse au traitement.

Nous nous sommes ainsi concentrés sur une cohorte d'enfants de plus de 6 ans traités avec la thérapie double VX-770/VX-809, suivis dans plusieurs centres pédiatriques, en France et en

Italie. Nous avons pris en considération 6 paramètres couramment mesurés chez les patients traités par modulateurs de CFTR et évoqués précédemment : VEMS, CVF, DEM25-75, IMC, nombre de jours d'antibiotiques, et concentration en ions chlorure dans la sueur, dont nous avons mesuré l'évolution durant les 6 premiers mois du traitement. A partir de ces mesures, nous avons identifié par ACP des paramètres indépendants permettant de caractériser la variabilité dans la réponse au traitement selon ces paramètres. L'ACP a aussi été utilisée comme méthode de réduction de dimension, avant de réaliser un groupement par la méthode k-means (décrite en introduction et à nouveau brièvement ci-après), permettant la détermination de groupes de réponse au traitement, à partir du nombre minimal de PCs permettant d'expliquer deux tiers de la variance totale de l'échantillon.

1.2 Matériels et méthodes

1.2.1 DESCRIPTION DES DONNEES

L'étude a été menée en tant qu'analyse rétrospective de données collectées durant le suivi de patients pédiatriques. Ces derniers étaient âgés de 6 à 18 ans, homozygotes pour la mutation F508del, traités avec la combo-thérapie VX-770/VX-809, et étaient suivis dans l'un des quatre centres suivants : l'Hôpital Necker Enfants Malades à Paris, les Hospices Civils de Lyon, le Centre Hospitalier Universitaire de Nantes, ou l'Ospedale pediatrico Bambino Gesù de Rome. Les patients étaient enrôlés avant de débiter le traitement par VX-770/VX-809 puis évalués à 6 mois selon les pratiques standardisées recommandées, assurant l'homogénéité des méthodes de mesure. Les 3 paramètres permettant d'évaluer les fonctions pulmonaires, et évoqués précédemment (VEMS, CVF, DEM25-75), ainsi que l'IMC, le nombre de jours d'antibiotiques dans les 6 derniers mois pour traiter les exacerbations pulmonaires à la fois en intraveineuse et par voie orale (ATB), et la concentration en ions chlorure dans la sueur (ST, pour Test de la Sueur), ont été mesurés au début du traitement (M0) et 6 mois plus tard (M6). Nous avons ainsi mesuré le changement dans ces paramètres entre M0 et M6, générant 6 nouvelles valeurs : $\Delta ppVEMS$, $\Delta ppCVF$, $\Delta ppDEM_{25-75}$, ΔIMC , ΔATB , ΔST où "Δ" signifie "Variation en absolue entre la référence et 6 mois de traitements". "pp" signifie "pourcentage prédit" et permet de normaliser les mesures par rapport à l'âge. Les mesures respiratoires ne sont ainsi pas données en valeur absolue, mais en pourcentage par rapport à la valeur moyenne des mesures obtenues pour les paramètres considérés chez des

enfants sains de même âge. De même, l'IMC n'est pas donné en valeur absolue, mais en z-score, une méthode permettant de normaliser la mesure par rapport à l'âge et le sexe du patient (une valeur de 0 signifie que l'enfant est exactement dans la moyenne d'IMC des enfants de même âge et de même sexe).

Tous les enfants ont été enrôlés entre 2017 et 2020, et tous les enfants enrôlés ont été conservés dans l'étude.

1.2.2 UTILISATION DE L'ACP

Comme évoqué en introduction, nous avons au cours de cette étude utilisé l'ACP pour décrire les données de notre étude (Jolliffe 1986). De cette manière, 6 nouvelles variables, les PC1 à PC6, ont été calculées comme combinaison linéaire des 6 paramètres définis ci-dessus et caractérisant initialement les patients. Ces nouvelles variables sont indépendantes et organisées de sorte que le pourcentage de la variance totale de l'échantillon capturé par les directions correspondantes décroisse de la PC1 à la PC6. En particulier, la PC1 correspond à la direction selon laquelle les patients sont le plus dispersés (voir l'explication détaillée sur l'ACP en introduction de cette thèse pour plus de détails).

Les PCs étant des combinaisons linéaires des paramètres initiaux, leurs coefficients associés permettent de déterminer lesquels participent le plus aux différentes PCs, et donc à la variabilité de l'échantillon. Ces coefficients sont contenus dans des matrices dites de « poids » : plus le poids d'un paramètre initial est élevé dans une PC, plus il contribue à cette dernière, et donc à la variance de l'échantillon selon la direction correspondante. Les poids peuvent aussi être assimilés à une forme de corrélation. Ainsi, des paramètres initiaux associés à des poids élevés dans la même PC sont corrélés, tandis qu'à l'inverse, des paramètres initiaux ayant des poids faibles dans la même PC sont essentiellement indépendants.

Les paramètres avec des poids élevés dans les premières PCs sont les plus importants pour séparer les patients selon leur réponse au traitement, à l'inverse des dernières PCs qui ne contiennent au contraire que peu, voire pas d'information, et peuvent donc être ignorées pour simplifier l'information globale et faciliter l'analyse et l'interprétation des résultats. Ainsi, dans notre étude, nous nous sommes concentrés sur le minimum de PCs nécessaires pour expliquer au minimum deux tiers de la variance globale de l'échantillon, permettant notamment de simplifier les représentations graphiques associées.

Afin de supprimer un biais dans le calcul de l'ACP vers les paramètres ayant des valeurs de base plus élevées, et donc des variations plus grandes, toutes les données ont été normalisées avant de réaliser l'ACP. De même, l'ACP ne pouvant pas fonctionner si le jeu de données considéré contient des données manquantes, une imputation de données a aussi été réalisée au préalable (la méthode détaillée sera décrite par la suite).

1.2.3 GROUPEMENT DES PATIENTS

Le groupement des patients a été réalisé après l'ACP, à partir des PCs expliquant le plus de variance uniquement. La méthode utilisée a été le groupement par K-means clustering (Lloyd 1957), qui permet de séparer les patients dans des groupes de variances égale, facilement descriptibles à partir des paramètres initiaux. Le nombre optimal de groupes a été déterminé par la méthode du C-index (Hubert 1976). Une fois obtenus, les groupes ont été comparés entre eux à l'aide des 6 paramètres initiaux, afin de mieux caractériser les profils de réponse associés à chaque groupe.

1.2.4 ANALYSES STATISTIQUES

Toutes les analyses ont été réalisées avec R (Ihaka 1996). Lors de réalisation de tests statistiques, une p-value inférieure à 0.05 était considérée comme statistiquement significative. Ces derniers n'étaient par ailleurs pas des tests paramétriques, étant donné que le présupposé de normalité des données ne pouvait pas être vérifié pour tous les paramètres. Les comparaisons entre patients, ainsi que les comparaisons entre groupes, ont donc été réalisées via des tests de rangs signés de Wilcoxon. Lorsque nécessaire (comparaison de valeurs de paramètres chez un même patient à M0 et à M6 notamment), des tests appariés ont été réalisés. Les corrélations entre variables ont quant à elles été réalisées par des tests de corrélation de Spearman. Enfin, des corrections de Benjamini-Hochberg ont été appliquées lorsque nécessaire pour prendre en compte l'effet des tests multiples.

1.2.5 IMPUTATION DES VALEURS MANQUANTES

51 données étaient manquantes sur la totalité de la cohorte. Bien que cela représentait moins de 4% des données totales, retirer tous les patients dont au moins une des valeurs était manquante aurait entraîné le retrait de 38% d'entre eux. Un algorithme d'imputation de données, MissForest

(Stekhoven 2012) a donc été utilisé. Il se présente sous la forme d'un package R et est basé sur un processus itératif à partir d'un modèle de random forest. Plusieurs autres algorithmes ont aussi été testés, notamment une imputation par PCA, par la méthode MICE, ou aléatoirement, mais MissForest était celui menant à la plus petite erreur de reconstruction.

Afin de s'assurer que les analyses ACP réalisées après imputation restaient statistiquement valables, deux méthodes différentes de ré-échantillonnage aléatoire ont été utilisées (Efron 1994), selon les recommandations dans le contexte d'imputations multiples pour l'ACP (Josse 2011). La première permettait de s'assurer que la petite taille de la cohorte (seulement 112 patients) n'avait pas d'impact sur l'ACP, tandis que la seconde permettait de confirmer la robustesse de l'ACP au processus d'imputation de valeurs manquantes.

Dans un premier temps, nous avons retiré 20% des patients de la cohorte avant de réaliser l'ACP sur les données restantes. Ce processus a été répété 100 fois afin de créer la figure 16 :

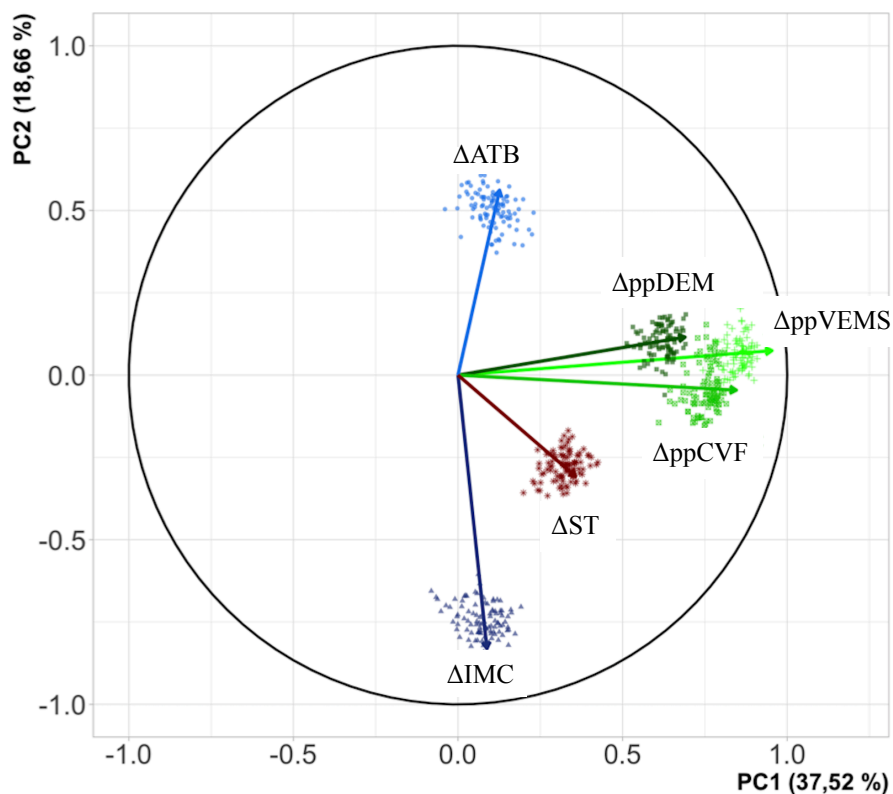


Figure 16 : Robustesse de l'ACP vis-à-vis de la taille de la cohorte. Chaque point coloré correspond à l'extrémité du vecteur correspondant pour un rééchantillonnage donné. La distribution des points décrit comment la taille et l'orientation de chaque paramètre varie du fait du rééchantillonnage dans le plan (PC1,PC2). Des points situés près de l'extrémité de la flèche correspondante sont caractéristiques d'une ACP robuste.

Dans cette figure, chaque point correspond à la participation d'un paramètre pour une itération donnée, dans les deux premières composantes de l'ACP. Chaque paramètre est donc décrit par 100 points, ainsi que par une flèche dont la pointe correspond à la participation du paramètre dans les deux premières composantes, dans le cas où toutes les données sont conservées. La distribution des points pour chaque paramètre décrit la variabilité du processus d'ACP du fait de la petite taille de la cohorte. Des points situés proches de la pointe de la flèche leur correspondant sont caractéristiques d'une ACP robuste, ce qui est effectivement le cas dans la figure.

Dans un second temps, 20% des données ont été retirées de manière aléatoire, sans retirer totalement des patients de la cohorte, avant de procéder à l'imputation de ces dernières, puis à la réalisation de l'ACP. Ce processus a lui aussi été répété 100 fois afin de créer la figure 17 :

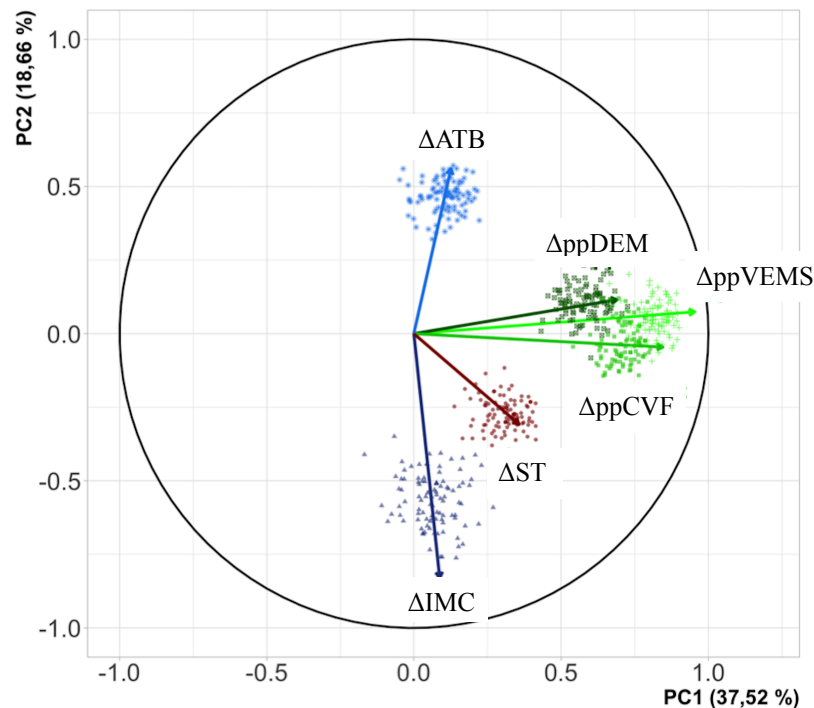


Figure 17 : Robustesse de l'ACP vis-à-vis de l'imputation des données manquantes. Chaque point coloré correspond à l'extrémité du vecteur correspondant pour une réimputation donnée. La distribution des points décrit comment la taille et l'orientation de chaque paramètre varie du fait des réimputations dans le plan (PC1,PC2). Des points situés près de l'extrémité de la flèche correspondante sont caractéristiques d'une ACP robuste.

L'interprétation de cette figure est la même que la précédente. Ici aussi, les points restent proches de la pointe de la flèche correspondante, ce qui témoigne de la robustesse de l'ACP à la présence de valeurs manquantes imputées par MissForest.

Ces deux analyses, considérées ensemble, démontrent que ni la petite taille de la cohorte, ni la variabilité induite par l'imputation de valeurs manquantes présentes dans le jeu de données original n'ont d'influence sur l'interprétation de l'ACP réalisée à partir de ces données.

1.3 Résultats

1.3.1 CARACTERISTIQUES DE LA COHORTE

Nous nous sommes dans un premier temps intéressés à la cohorte de patients dans son ensemble, sans réaliser d'ACP ni de groupement. Toutes les caractéristiques de ces 112 patients homozygotes pour la mutation F508del sont présentes dans le tableau en annexe 1. Les patients n'ont que de très légers symptômes pulmonaires, comme le montre la valeur moyenne du ppVEMS à 87% à M0 au niveau global. De manière similaire, la plupart des patients présentent un statut nutritionnel normal avec notamment un Z-score d'IMC moyen de seulement -0.6 à M0.

Néanmoins, les 24 patients âgés de 6 à 12 ans apparaissent comme ayant un statut respiratoire significativement meilleur que celui des 88 patients âgés de 12 à 18 ans, comme le montre un ppVEMS et un ppDEM₂₅₋₇₅ significativement plus élevés ($p < 10e-3$ et $p = 0.01$ respectivement). A partir de ce constat, et pour pouvoir étudier plus spécifiquement le cas d'une cohorte sans symptômes respiratoires, nous avons étudié les 6-12 ans séparément des 12-18 ans.

De manière générale, on constate dans les deux cohortes une amélioration significative de l'IMC et du ST après 6 mois de traitement par VX-770/VX-809. A l'inverse, ni le ppVEMS, ni le ppDEM₂₅₋₇₅, n'évoluent de manière significative en 6 mois.

Notons aussi que de nombreuses valeurs étaient manquantes : 42 dans la cohorte des 12-18 ans (4.0% des données) et 9 dans celle des 6-12 ans (3.1% des données), la plupart étant liées au ST. Retirer tous les patients présentant au moins une valeur manquante aurait mené à la suppression de 43 patients avant de pouvoir réaliser l'analyse ACP. Une imputation des valeurs manquantes a donc été réalisée au préalable, telle que décrite précédemment. Cette dernière n'a pas changé l'interprétation de nos données à l'échelle de la cohorte globale, et plusieurs analyses de robustesse de l'imputation ont été réalisées, témoignant de son utilité.

1.3.2 ANALYSE DES DONNEES DES PATIENTS AGES DE 12 A 18 ANS

Nous nous sommes ensuite concentrés sur les données de patients âgés de 12 à 18 ans. L'analyse ACP a déterminé 6 PCs, notées PC1 à PC6. La quantité de variance expliquée par chaque composante est décrite dans la table en annexe 2, tandis que la table des poids est détaillée dans la table en annexe 3. Comme expliqué précédemment, nous n'avons retenu que les 3 premières PCs dans nos analyses (PC1 PC2, PC3), car elles permettaient à elles seules de résumer plus de 70% de la variance totale de notre échantillon (avec respectivement 38%, 19% et 17% de la variance totale expliquée).

La figure 18 représente la participation des paramètres cliniques dans les trois premières composantes de l'ACP, sous la forme d'un cercle de corrélation. A gauche, les PC1 et PC2 sont représentées, tandis qu'à droite, ce sont les PC2 et PC3. Dans cette figure, chaque paramètre est représenté par une flèche dont l'abscisse et l'ordonnée de la pointe correspondent à leur corrélation avec les PCs correspondantes. De plus, le pourcentage de variance de l'échantillon expliqué par chaque PC est représenté le long de l'axe correspondant.

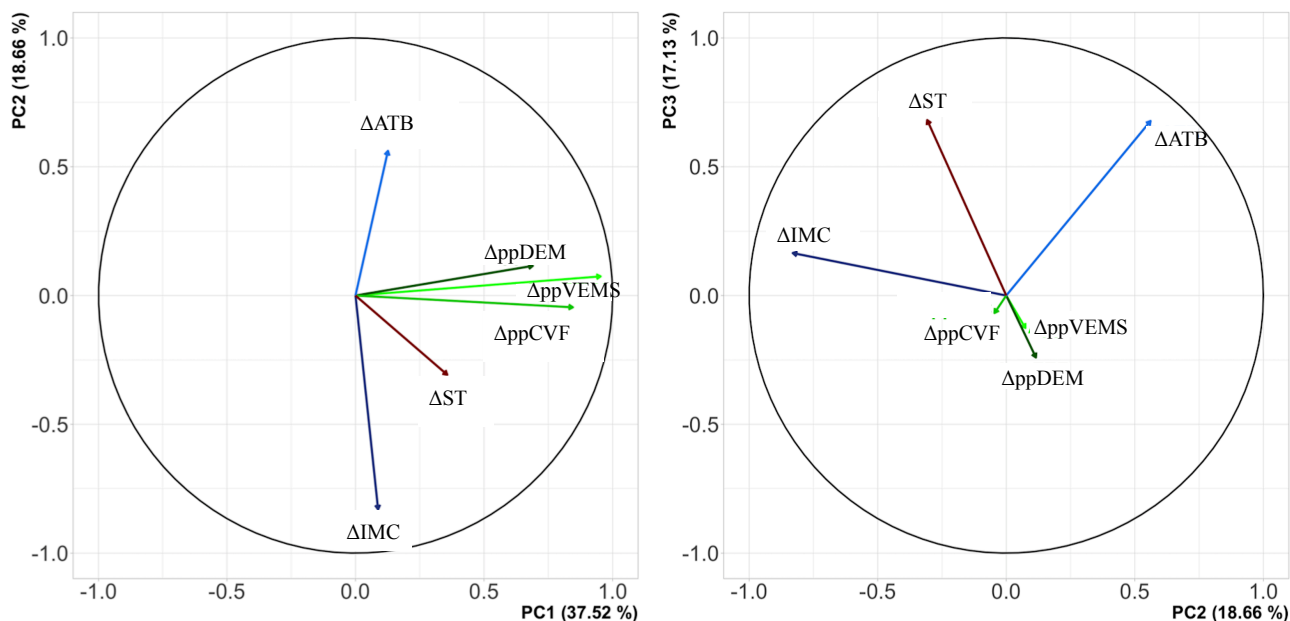


Figure 18 : Cercle de corrélation entre les paramètres cliniques et les trois premières composantes de l'ACP pour les patients de 12-18 ans. (A) Représentation graphique des PC1 et PC2. (B) Représentation graphique des PC2 et PC3. Chaque paramètre est représenté par un vecteur dont l'abscisse et l'ordonnée marquent sa corrélation avec les composantes de l'ACP. Le % de la variance globale expliquée par chaque PC est indiqué au niveau des labels.

L'étude de cette figure, ainsi que de la table en annexe 2, montre que l'évolution des trois paramètres respiratoires (ΔppVEMS , ΔppCVF , $\Delta\text{ppDEM}_{25-75}$) sont les contributeurs les plus importants à la PC1. Cependant, les valeurs de poids similaires pour ces trois paramètres dans la PC1 témoignent de la corrélation entre ces derniers, et donc de la redondance des informations qu'ils portent. La même étude des poids montre que le ΔST est aussi corrélé à la PC1, mais à un degré moindre, ce qui est aussi visible dans la figure précédente pour laquelle la flèche de ΔST est moins bien alignée avec l'axe de la PC1 que celles des trois paramètres respiratoires.

Ces poids montrent que le ΔATB et le ΔIMC sont tous les deux fortement corrélés à la PC2. De plus, comme attendu, ces deux paramètres pointent dans des directions opposées dans la figure précédente : lorsque l'IMC d'un patient augmente, la quantité d'antibiotiques prescrits pour ce dernier au cours des 6 derniers mois a tendance à diminuer. Enfin, le ΔST et le ΔATB sont les deux contributeurs les plus importants à la PC3.

Un groupement a ensuite été réalisé pour déterminer différents profils de réponse au traitement par VX-770/VX-809. Le critère de C index a permis de déterminer que le nombre idéal de groupes était de 4 dans notre étude. Les figures suivantes montrent la répartition des patients au

sein des différents groupes dans l'espace généré par les 2 premières (figure 19) et les 3 premières (figure 20) composantes de l'ACP, ainsi que la distribution des paramètres pour chaque groupe (figure 21).

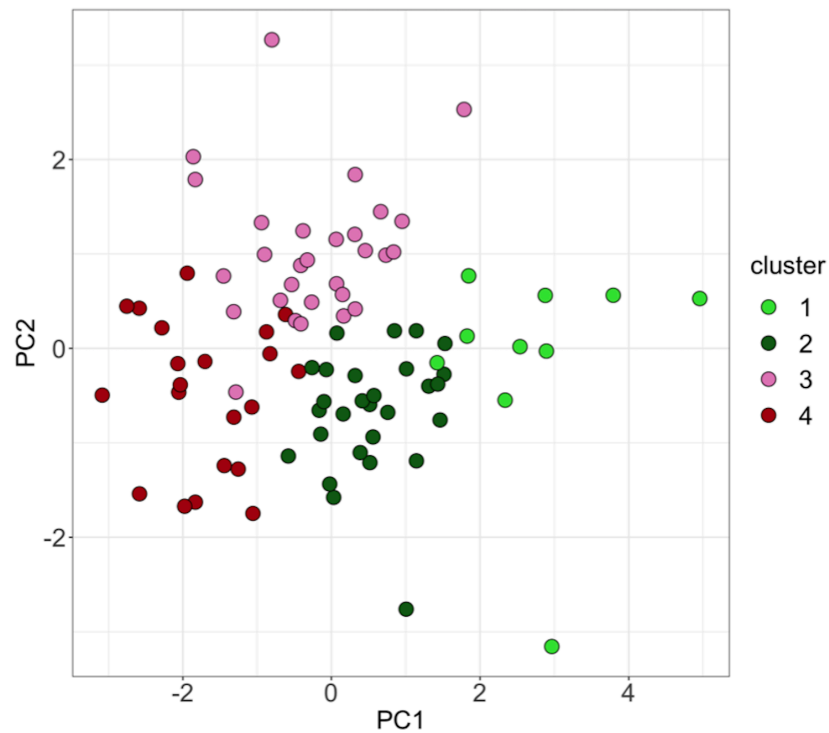


Figure 19 : Projection des patients de 12-18 ans dans l'espace PC1-PC2. Chaque point correspond à un patient et chaque couleur correspond à un groupe (cluster). Les coordonnées de chaque patient ont été obtenues par combinaison linéaire des paramètres contribuant aux PCs.

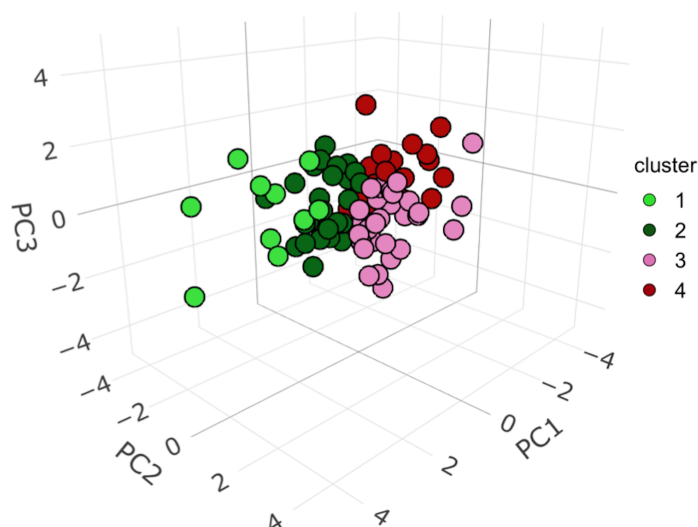


Figure 20 : Projection des patients de 12-18 ans dans l'espace PC1-PC2-PC3. Chaque point correspond à un patient et chaque couleur correspond à un groupe (cluster). Les coordonnées de chaque patient ont été obtenues par combinaison linéaire des paramètres contribuant aux PCs.

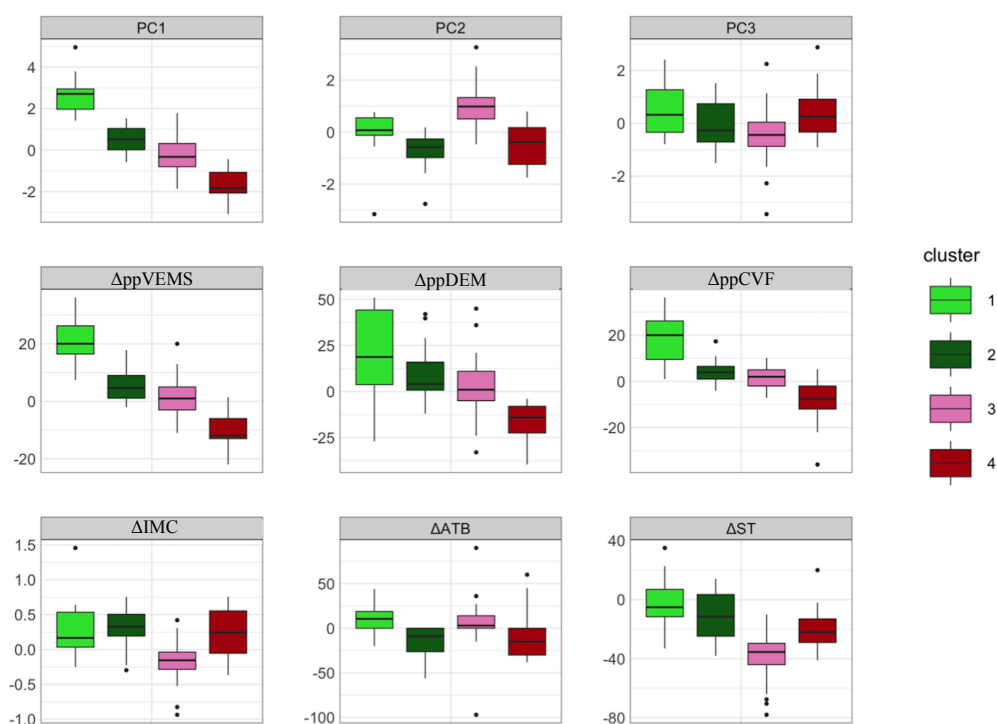


Figure 21 : Distribution des paramètres dans les 4 groupes (clusters) pour les patients de 12-18 ans. Les valeurs pour les PC1, PC2 et PC3 ont été obtenues en par combinaison linéaire des paramètres contribuant aux PCs.

La table en annexe 4 montre les valeurs de la figure 21 sous forme de tableau, et compare les groupes selon les paramètres cliniques pris en compte au cours de cette étude.

On note ainsi que les groupes 1 et 4 se caractérisent par l'évolution de leurs paramètres respiratoires. En particulier, le groupe 1 se caractérise par une amélioration des trois paramètres $\Delta ppVEMS$, $\Delta ppCVF$ et $\Delta ppDEM_{25-75}$. L'amélioration du VEMS est ainsi significativement plus importante pour les patients appartenant au groupe 1 que pour tous les autres groupes (amélioration de ce paramètre de 21 points en moyenne). A l'inverse, la diminution de ce même paramètre est significativement plus importante au sein du groupe 4 que dans tout autre groupe (baisse de 10 points en moyenne). Un pattern similaire est observable pour les paramètres CVF et DEM_{25-75} , mais à un degré moindre que pour le VEMS. En particulier, on note que le $\Delta ppDEM_{25-75}$ n'est pas significativement supérieur dans le groupe 1 par rapport aux groupes 2 et 3. La combinaison des 3 paramètres respiratoires est résumée dans le paramètre PC1 de la figure 21. Ce dernier ne permet pas d'améliorer la séparation entre les groupes, ce qui soutient l'idée que ces paramètres portent des informations redondantes. De manière intéressante, les patients du groupe 1 présentent un $ppCVF$ significativement plus faible que celui de tous les autres groupes au moment du début du traitement, ainsi qu'un $ppVEMS$ plus faible que celui des patients du groupe 4. En outre, les patients du groupe 1 présentent la plus petite baisse de ST (baisse de seulement 1.6 en moyenne).

Les patients des groupes 2 et 3 présentent des évolutions intermédiaires en termes de fonctions respiratoires, et se différencient particulièrement par leurs évolutions en ATB, en IMC et en ST. Les patients du groupe 2 présentent ainsi la plus grande baisse en ATB, tandis que ceux dans le groupe 3 ont la plus baisse importante en ST, et, au contraire de tous les autres groupes, voient leur IMC baisser en moyenne. Notons aussi que les patients du groupe 3 présentent un ST significativement plus élevé que celui de n'importe quel autre groupe au moment du début du traitement.

1.3.3 ANALYSE DES DONNEES DES PATIENTS AGES DE 6 A 12 ANS

Pour le cas de la cohorte des patients âgés de moins de 12 ans, comme le montre la table en annexe 5, les deux premières composantes de l'ACP résument à elles seules plus de 66% de la variance totale de la cohorte. Nous nous sommes donc concentrés sur les deux premières PCs pour l'analyse de ces données. De plus, étant donné le faible nombre de patients (24), nous n'avons pas réalisé de groupement.

De manière similaire à ce que l'on a pu observer chez les patients âgés de 12 à 18 ans, la PC1 est essentiellement corrélée aux trois paramètres respiratoires, comme le montrent la table en annexe 5 ainsi que la figure 22 suivante, représentant un cercle de corrélation similaire à ceux établis pour les patients de 12 à 18 ans :

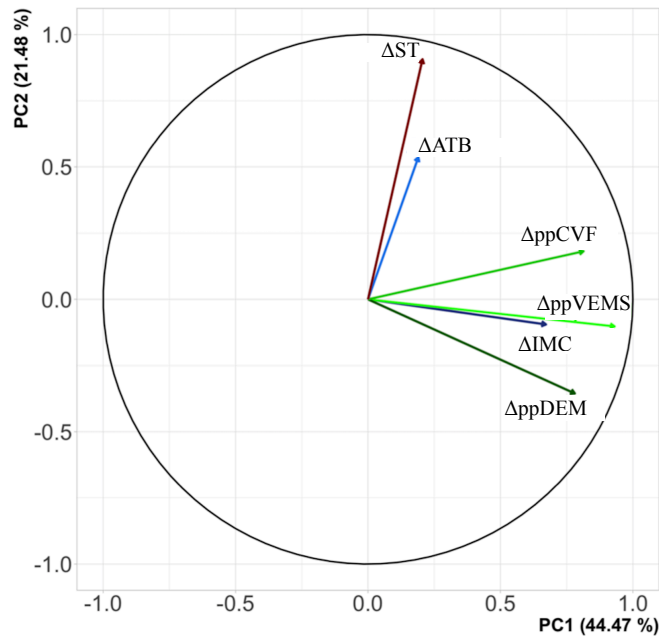


Figure 22 : Cercle de corrélation entre les paramètres cliniques et les deux premières composantes de l'ACP pour les patients de 6-12 ans. Chaque paramètre est représenté par un vecteur dont l'abscisse et l'ordonnée marquent sa corrélation avec les composantes de l'ACP. Le % de la variance globale expliquée par chaque PC est indiqué au niveau des labels.

Les poids correspondants sont détaillés dans la table en annexe 6. On note que les 3 paramètres respiratoires restent très corrélés entre eux et portent donc eux aussi des informations redondantes. Cependant, contrairement à la cohorte des patients de plus de 12 ans, ici, le ΔIMC est aussi un contributeur important à la PC1. Enfin, les ΔST et ΔATB sont les deux contributeurs principaux à la PC2.

1.4 Discussion des résultats

Cette étude visait à caractériser l'hétérogénéité de la réponse à VX-770/VX-809 chez des patients homozygotes pour la mutation F508del, et à déterminer différents profils de réponse à ce

traitement. Pour cela nous avons utilisé des méthodes mathématiques basées sur l'ACP pour identifier des variables non redondantes, le groupement par k-means pour caractériser des groupes de réponses, et un algorithme de random forest pour imputer les données manquantes.

Cette étude a également mis en lumière la manière donc les paramètres étudiés évoluent les uns par rapport aux autres sur une période de six mois à partir du début du traitement, mettant ainsi en évidence différents types de réponse à ce dernier.

De manière intéressante, le paramètre nutritionnel, évalué dans notre étude par la mesure de l'IMC, est apparu comme étant important pour caractériser la réponse chez les patients âgés de 6 à 12 ans, alors que ce n'était pas le cas dans la cohorte des 12-18 ans.

Compte tenu de l'efficacité des modulateurs de CFTR, de plus en plus d'enfants vont bénéficier de ces thérapies dans les années à venir, notamment car si elles sont mises en œuvre assez tôt dans la vie des patients, elles pourraient ralentir l'évolution de la maladie (Accurso 2018). Cependant, le fait que les enfants présentent des symptômes très légers, voire soient asymptomatiques, rend l'évaluation dans des études en vie réelle très difficile (Stanojevic 2016). Comme expliqué précédemment, des paramètres tels que le LCI sont actuellement testés pour déterminer s'ils peuvent témoigner de la réponse à un traitement (Ratjen 2018). Cependant, la mesure de ces nouveaux paramètres est complexe et ne peut pas être réalisée dans tous les centres. Nous avons cherché à savoir si les différents paramètres mesurés de manière routinière et standardisée dans tous les centres européens permettaient de déterminer différents types de réponse. Contrairement aux études déjà menées où l'effet d'un traitement était évalué en terme de réponse moyenne sur la cohorte entière, nous nous sommes concentrés sur l'hétérogénéité de la réponse dans le cadre d'un suivi en vie réelle. Nous avons utilisé l'ACP pour réduire nos 6 paramètres d'étude à un plus petit nombre de composantes principales, et ce, sans perdre trop d'information, avant de réaliser un groupement. Nous avons en outre évalué la redondance entre nos paramètres d'étude pour déterminer un ensemble minimal de paramètres nécessaires dans l'optique d'évaluer la réponse des patients pédiatriques aux traitements par modulateurs. Par conséquent, notre analyse diffère de celles effectuées habituellement puisque nous souhaitons déterminer quels paramètres varient de manière similaire en réponse à un traitement. Nous avons non seulement évalué quels facteurs étaient impliqués dans l'hétérogénéité de la réponse au sein de la cohorte, mais nous avons également cherché à trouver quelle combinaison de paramètres permettait d'expliquer le plus de variabilité les données cliniques considérées. De plus, grâce au groupement, nous avons pu définir différents profils de réponse. A notre connaissance, une seule autre étude s'est intéressée à l'hétérogénéité de la réponse à un traitement chez des patients atteints de bronchectasie, et a

également conclu que la réponse était très hétérogène (Sibila 2021). Ces observations renforcent la nécessité de comprendre comment de multiples facteurs participent à une réponse à un traitement.

Il est important de noter que la méthodologie que nous proposons dans cette étude peut être appliquée à d'autres études. En particulier, suite à l'avènement de la trithérapie pour traiter les patients atteints par la mucoviscidose (Zemanick 2021), il serait intéressant d'appliquer cette méthodologie à une cohorte de patients traités par VX-770/VX-661/VX-445.

La principale limite de notre étude est le petit nombre de patients. Nous avons donc vérifié la solidité statistique de nos résultats par une approche de ré-échantillonnage aléatoire, en évaluant comment les résultats de l'ACP auraient changé après la suppression d'une partie de nos données. En outre, il y avait des données manquantes dans notre cohorte. Bien que seuls quelques résultats manquaient (environ 4 %), ignorer les patients avec des données manquantes aurait conduit à écarter environ 38 % des patients, ce qui est inadmissible dans des cohortes déjà petites. Nous avons donc mis en œuvre une méthode d'imputation des données manquantes, et avons évalué si les résultats de l'ACP pouvaient changer du fait d'erreurs potentielles lors de l'imputation de ces données. Nous avons ainsi pu vérifier que les résultats de l'ACP étaient stables à la fois vis-à-vis du ré-échantillonnage et des potentielles erreurs d'imputation, comme décrit en détail dans la partie « imputation des valeurs manquantes » de cette section.

Une autre limite est l'absence de suivi au-delà de 6 mois. Bien que cette période soit similaire à celle utilisée dans les études cliniques (Ratjen 2017, Walker 2019) et permette de mettre en évidence une hétérogénéité significative dans la réponse au traitement, l'effet possible à long terme du traitement et l'hétérogénéité à long terme ne peuvent pas être évalués par cette seule étude.

Le groupement des patients a permis de mettre en évidence deux niveaux de réponse dans la cohorte des 12-18 ans. Premièrement, et même si elle n'était pas significative au niveau de l'échantillon entier, l'évolution du VEMS a permis de caractériser les " bons répondeurs respiratoires " avec une amélioration significative (le groupe 1), les " mauvais répondeurs respiratoires " avec une détérioration significative (le groupe 4) et les groupes avec un changement minimal, c'est-à-dire qui reste dans le cadre de la variabilité intrinsèque à ce test respiratoire (Cooper 1990). Cela montre que l'évolution du VEMS est d'une importance cruciale pour discriminer la réponse des patients pédiatriques âgés de 12 à 18 ans traités par VX-770/VX-809. En outre, ni le Δ ppCVF, ni le Δ ppDEM₂₅₋₇₅, n'ont amélioré la caractérisation de ces groupes, ce qui suggère qu'ils sont redondants avec le Δ ppVEMS et n'apportent pas d'informations supplémentaires. Il est intéressant de noter que les patients du groupe 1 (ceux qui se sont le plus améliorés) avaient le plus mauvais ppVEMS au début du traitement, de même qu'un ppCVF significativement plus faible que celui des autres

groupes. Cela suggère que dans une population pédiatrique, les patients ayant des paramètres de fonctions respiratoires plus faibles, et en particulier un VEMS plus faible, ont un potentiel d'amélioration plus élevé. Ceci est soutenu par des données similaires chez les adolescents, qui ont montré une amélioration du VEMS chez ceux dont le VEMS de départ était le plus faible (Wainwright 2015), alors qu'aucune amélioration du VEMS n'a été détectée chez des patients dont le VEMS était normal (Sagel 2021, Milla 2017). Dans l'ensemble, l'évolution du VEMS dans cette cohorte souligne l'importance de considérer des sous-groupes de patients. En effet, la prise en compte de l'évolution moyenne du VEMS au niveau de l'échantillon global uniquement aurait conduit à la conclusion qu'il n'y a pas d'amélioration significative. Au contraire, l'analyse des sous-groupes met en évidence une amélioration significative dans des sous-populations spécifiques de la cohorte et souligne l'hétérogénéité de la réponse au traitement. Les patients qui ont le plus bénéficié du traitement sont ceux qui présentaient un ppCVF significativement plus faible au début du traitement par rapport à tout autre groupe, et un ppVEMS significativement plus faible par rapport au groupe 4.

Pour caractériser davantage la réponse des patients n'ayant pas de changement significatif en VEMS (groupes 2 et 3), un deuxième niveau de caractérisation de la réponse a été défini par la PC2 et comprend l'évolution du ST, de l'IMC et des ATB. Le groupe 2 était ainsi caractérisé par une augmentation de l'IMC et une diminution de la consommation d'ATB (donc par moins d'exacerbations pulmonaires). Cela met en évidence un profil spécifique de patients présentant une amélioration concomitante de leur état nutritionnel et de leur niveau d'infection, et ce malgré un changement minime du ppVEMS. L'importance de la mesure du nombre de jours d'antibiotiques en tant que facteur indépendant caractéristique de la réponse à un traitement dans les cohortes pédiatriques avait également été mise en évidence dans d'autres cohortes avec des patients présentant un ppVEMS normal au début du traitement (Burgel 2020, Stick 2008, Britto 2002). Cette diminution du recours aux antibiotiques peut refléter une amélioration du niveau d'inflammation pulmonaire et de l'immunité innée, basée sur une amélioration de l'activité antimicrobienne et de la fonction des cellules immunitaires (Khan 2019, Geitani 2020). Cela souligne par ailleurs le bénéfice potentiel de ces thérapies, même chez les patients dont la fonction pulmonaire est normale. Cet aspect est crucial dans la population pédiatrique qui présente une maladie sans symptômes apparents dans une perspective de soins préventifs.

De manière similaire aux résultats obtenus dans notre cohorte, une amélioration significative de l'IMC a déjà été rapportée chez des enfants de 12 à 18 ans traités par VX-770/VX-809 (Sagel 2021, Burgel 2020, Wainwright 2015). Cela peut refléter non seulement une augmentation de

l'apport calorique, mais aussi une amélioration du statut pancréatique, et peut-être un effet systémique d'un anabolisme accru dû à une diminution de l'état inflammatoire (Barry 2014). De manière inattendue, nous avons également identifié un groupe de patients dont l'IMC s'est aggravé (groupe 3). Nous n'avons pas d'explication claire à ce sujet, mais le fait que ces patients n'aient présenté aucune amélioration du VEMS ou des ATB montre que le traitement par modulateurs n'a pas du tout amélioré la condition de ces patients sur le plan clinique. Il est intéressant de noter que ces patients sont aussi ceux associés à la diminution de ST la plus importante, alors qu'à l'inverse, les patients présentant la meilleure amélioration respiratoire sont aussi ceux ayant la baisse de ST la plus faible. Cela démontre que, même si la concentration en ions chlorure dans la sueur diminue au niveau de la cohorte globale, son évolution n'est pas indicative d'un bénéfice clinique au niveau individuel dans la population pédiatrique. Ceci est en accord avec des études antérieures chez les adultes montrant que la concentration en ions chlorure dans la sueur diminue en réponse au traitement avec des modulateurs de CFTR, mais sans qu'il n'y ait de corrélation avec l'évolution du ppVEMS (Barry 2014, Durmowicz 2013, Masson 2019). Par ailleurs, les patients du groupe s'étant le plus amélioré d'un point de vue du ST présentaient une valeur de ce paramètre significativement plus élevée que dans tout autre groupe à M0, et il est donc possible qu'ils aient eu une plus grande marge d'amélioration pour ce paramètre.

La cohorte des 6-12 ans a été analysée séparément, car les patients de cette dernière avaient une fonction respiratoire significativement meilleure que celle des enfants de 12-18 ans au début du traitement. Dans cette cohorte plus jeune, le ppVEMS ne s'est pas amélioré de manière significative, comme cela a déjà été rapporté dans des études menées avec le VX-770/VX-809 et le VX-661/VX-770 dans un groupe d'âge similaire (6-11 ans) (Ratjen 2017, Walker 2019, Milla 2017). Mais de manière plus importante, le profil de l'ACP dans cette cohorte était différent de celui de la cohorte des 12-18 ans : la PC1 était non seulement corrélée à la variation du VEMS, mais aussi à celle de l'IMC. Cela montre que dans cet échantillon plus jeune avec une fonction respiratoire normale, l'IMC capte une partie de la réponse et constitue un paramètre important. Ceci est indirectement confirmé par le fait que dans la plupart des études menées dans ce groupe d'âge, l'IMC est systématiquement amélioré, contrairement au VEMS (Ratjen 2017, Davies 2013). L'évolution de l'IMC est généralement considérée comme un endpoint secondaire. Ici, nous montrons que l'évolution de l'IMC est un critère important chez les jeunes patients pour distinguer les répondeurs des non-répondeurs.

1.5 Conclusion de l'étude

Dans cette étude, nous proposons un cadre mathématique global adapté aux études de la vie réelle qui pourrait être transposé à d'autres traitements actuels.

Dans le cas de la combinaison VX-770/VX-809 chez des enfants de 6-18 ans, il a permis de montrer que les variations du VEMS, de l'IMC et du nombre de jours d'ATB expliquent la majeure partie de l'hétérogénéité de la réponse des enfants. En particulier, l'IMC revêt une importance toute particulière dans l'hétérogénéité de la réponse chez les enfants de 6 à 12 ans dont la fonction respiratoire est normale. Nous montrons aussi que les améliorations de la fonction pulmonaire, de l'IMC, de la fréquence des exacerbations et du ST sont dissociées chez les enfants atteints par la mucoviscidose.

Néanmoins, cette étude souligne également l'impossibilité d'établir un score unique clair, qui serait lié à des facteurs biologiques et qui permettrait d'établir clairement l'efficacité ou non d'un traitement chez les patients atteints de mucoviscidose, ou tout du moins les paramètres cliniques mesurés en routine apparaissent comme insuffisants pour atteindre cet objectif. Il est donc toujours nécessaire de rechercher des facteurs de réponse clairs, facilement mesurables, et dont l'évolution en réponse à un traitement par modulateurs serait homogène chez tous les patients porteurs des mêmes mutations. Dans cette optique, l'étude des composants de la sueur apparaît comme un élément intéressant : facilement collectable, elle n'a cependant été que peu étudiée. Dans ma seconde étude, je me suis donc intéressé à l'étude de la protéomique de la sueur de patients traités par modulateurs.

2. Analyse protéomique de la sueur de patients

L'article associé à cette section est publié dans Skin Health and Disease et donné en annexe de ce manuscrit.

2.1 Introduction

L'une des caractéristiques de la mucoviscidose est une concentration élevée d'ions chlorures dans la sueur en raison de l'absence d'une protéine CFTR fonctionnelle dans les glandes sudoripares (Elborn 2016). Malgré cette particularité, la sueur de patients atteints par la mucoviscidose n'a encore jamais été prise en considération pour des études omiques. Pourtant, une telle étude pourrait présenter un intérêt double.

D'une part, une analyse poussée de la protéomique de la sueur de patients atteints par la mucoviscidose, et traités par modulateurs, comparée à celle de patients non traités, pourrait mettre en évidence des biomarqueurs d'intérêt dont l'évolution au cours du traitement serait associée à différents aspects de la réponse du patient à ces traitements. S'il est possible d'identifier de tels biomarqueurs dans la sueur, ils pourraient être utilisés comme l'un des paramètres de référence pour évaluer l'efficacité de nouveaux traitements lors d'essais cliniques.

D'autre part, une telle analyse différentielle entre des patients atteints par la mucoviscidose et des contrôles sains pourrait permettre de mieux comprendre la physiopathologie de la maladie, en particulier au niveau de la peau. En effet, les patients atteints par la mucoviscidose présentent une KPA, caractérisée par un plissement important de la peau qui apparaît rapidement après une brève immersion dans l'eau, et associée à des sensations de picotement et de douleur (Alexopoulos 2021). L'importance de ce symptôme est sûrement sous-estimée, alors que la KPA est observée chez 41 à 84 % des patients atteints de mucoviscidose suivant les études. Ce symptôme reste peu étudié, ce qui peut expliquer cet intervalle important (Arkin 2012, Berk 2009).

La physiopathologie de la KPA et le lien de cette dernière avec la dérégulation de CFTR ne sont pas encore totalement élucidés. Des études récentes ont montré que ce symptôme était diminué par les modulateurs de CFTR chez les patients atteints par la mucoviscidose, ce qui établit un lien entre la KPA et la dérégulation de CFTR (Jacobi 2022).

Afin de rechercher de nouveaux biomarqueurs d'intérêt d'une part, mais d'autre part afin de mieux comprendre la physiopathologie de la KPA chez les patients atteints par la mucoviscidose, nous avons effectué une analyse approfondie du protéome de la sueur de patients atteints par la

maladie, traités ou non par des modulateurs de CFTR, et nous les avons comparés à celui de personnes saines. Dans un premier temps, les traitements considérés n'incluaient pas la récente trithérapie VX-770/VX-661/VX-445. Les résultats associés à cette trithérapie seront présentés à la fin du chapitre.

2.2 Matériels et méthodes

2.2.1 CARACTERISATION DES PATIENTS

Pour cette étude nous avons initialement recruté 78 personnes, comprenant 25 patients atteints par la mucoviscidose et non traités par modulateurs, 25 patients traités par différents modulateurs de CFTR, 16 individus sains porteurs d'une mutation à l'origine de la mucoviscidose et d'un allèle non muté (appelés hétérozygotes par la suite), et 12 individus sains sans allèle muté. Le génotype des patients a été classé selon le mécanisme génétique du variant (classe 1 : non-sens, délétion ou épissage sévère ; classe 2 : mauvais repliement ; classe 3 : défaut d'activation ; classe 4 : défaut de conductance ; classe 5 : épissage avec fonction résiduelle. Se reporter à l'introduction de ce manuscrit pour plus de détails sur les classes de mutations). Tous les participants ont au préalable signé un formulaire de consentement éclairé pour que leur sueur soit collectée et analysée.

L'état clinique des patients lors de leur visite d'inclusion a été évalué par la mesure des paramètres cliniques habituels évoqués dans la section précédente, incluant notamment une évaluation du statut pancréatique, et du VEMS, exprimé en pourcentage prédit (ppVEMS). La microbiologie des crachats a aussi été évaluée afin de s'assurer qu'aucun patient n'était surinfecté au moment du prélèvement. L'induction de sueur a été réalisée par iontophorèse de pilocarpine, selon la méthode décrite dans la littérature (Sermet Gaudelus 2021). La sueur a ensuite été recueillie à l'aide du Macroduct Wescor (ELITech, Puteaux, France) et la concentration en ions chlorure a été évaluée par colorimétrie (chloridomètre 926S Sherwood, Servilab, Le Mans, France). L'analyse protéomique a ensuite été réalisée à partir de vingt microlitres de l'échantillon de sueur collecté.

Nous nous sommes concentrés sur les 25 patients atteints par la mucoviscidose et non traités au moment du prélèvement, âgés en moyenne de 9,1 ans, tous porteurs de deux mutations de CFTR causant la mucoviscidose (11 homozygotes F508del; 7 hétérozygotes porteurs de F508del ; 7 porteurs de mutations non-sens, d'épissage ou de grandes délétions). La table en annexe 7 récapitule la liste de ces patients.

Le patient 13 était porteur d'une mutation faux-sens rare, p.Phe191Val (F191V), ainsi que de F508del et présentait une concentration normale en ions chlorure dans la sueur avec seulement 22 mmol/L. Cependant, le taux de sécrétion de sueur β -adrénergique était dans la gamme de celui des patients atteints par la mucoviscidose, ce qui laisse à penser que ce génotype est associé à une activité réduite de CFTR (Nguyen-Khoa 2022).

Le patient 20 portait la mutation d'épissage 2789+5G>A, connue pour être associée à la production de transcrits normalement épissés, mais en quantité réduite (Duguépéroux 2005).

Aucun de ces patients n'était sous traitement au moment du prélèvement de leur sueur. Trois patients (patients 13, 20, 25) ne présentaient pas de défaut au niveau pancréatique. Les patients 2, 3 et 21 présentaient une colonisation chronique de leurs expectorations par *Pseudomonas.aeruginosa* ou *Alcaligenes.spp* et étaient traités par antibiotiques inhalés.

Les 12 individus sains sans allèles muté (notés WT dans la table 5) avaient un âge moyen de 12,5 ans et une concentration normale en ions chlorure dans leur sueur.

Nous avons en outre enrôlé 16 individus sains hétérozygotes, et 25 patients atteints par la mucoviscidose et sous traitement par des modulateurs de CFTR au moment du prélèvement. 3 l'étaient par VX-770 (respectivement depuis 1.5, 2.5 et 5 ans) et 22 par VX-770/VX-809, mais avec des durées de traitement très variables (de 4 mois à 4 ans).

2.2.2 PREPARATION DES ECHANTILLONS EN VUE D'UNE ANALYSE PAR SPECTROMETRIE DE MASSE

Pour chaque échantillon, 20 μ l de sueur collectés sont traités dans une micro-colonne S-Trap dont le fonctionnement a été brièvement expliqué en introduction de ce manuscrit. Au préalable, les protéines sont séparées selon leur masse moléculaire par migration dans un gel comprenant l'agent dénaturant sodium dodecyl sulfate (SDS), puis réduites et alkylées à température ambiante. Les mélanges ont ensuite été chargés sur les colonnes S-Trap. Six étapes de lavage sont effectuées pour éliminer le SDS, puis les échantillons sont digérés avec de la trypsine (à environ 50°C). Enfin, après élution, les peptides sont séchés sous vide puis resuspendus dans de l'eau de qualité HPLC (pour high performance liquid chromatography, de l'eau adaptée à une analyse en chromatographie en phase liquide) avant analyse par chromatographie en phase liquide et spectrométrie de masse. La concentration en peptides des échantillons obtenus est déterminée en utilisant un Nanodrop 2000 de Labtech France, outil dédié à la réalisation de cette mesure à partir de quelques microlitres d'un échantillon.

2.2.3 ANALYSE CHROMATOGRAPHIE/SPECTROMETRIE DE MASSE

Pour chaque échantillon, 400ng de peptides ont été injectés sur un système HPLC (chromatographie liquide à haute performance) nanoelute (Bruker Daltonics, Allemagne) couplé à un spectromètre de masse timsTOF Pro (Bruker Daltonics, Allemagne). La séparation des peptides par élution sur la colonne d'HPLC est réalisée avec un gradient croissant de solvant organique. Ainsi, les peptides les plus hydrophiles migrent le plus rapidement, tandis que les plus hydrophobes sont élués plus tardivement, lorsque la teneur en solvant organique devient plus élevée. Les données de spectrométrie de masse sont acquises par la méthode dite parallel accumulation serial fragmentation, permettant d'analyser plusieurs protéines à la fois et donc d'accroître la quantité d'information obtenue par rapport à une acquisition séquentielle standard.

2.2.4 ANALYSE DES DONNEES

Le spectromètre de masse délivre pour chaque échantillon des données brutes comprenant pour chaque peptide la valeur m/z (ratio masse sur charge), l'intensité, et le temps de rétention dans la colonne de chromatographie en phase liquide (liée au caractère hydrophile ou hydrophobe du peptide). A partir de bases de données de protéines, des logiciels dédiés identifient alors les peptides et les protéines associées qui étaient présents dans l'échantillon. Dans notre cas, nous avons utilisé le logiciel MaxQuant, et recherché les protéines avec le moteur de recherche Andromeda dans la base de données UniProtKB/Swiss-Prot Homo sapiens (un peu plus de 20000 entrées répertoriées). Nous avons imposé un grand niveau de précision pour l'identification, en autorisant une erreur de masse mesurée par le spectromètre pour chaque peptide de seulement 3ppm. De plus, seuls les peptides d'au moins 7 acides aminés de long ont été considérés. Pour prendre en compte les effets de modifications chimiques des protéines en amont de leur analyse par spectrométrie (réduction et alkylation), la carbamidométhylation des cystéines, l'oxydation des méthionines, ainsi que l'acétylation N-terminale ont été prises en compte en tant que modifications possibles lors de la recherche dans la base de données de protéines. En outre, les peptides associés à des contaminants connus (qui seraient issus d'un des agents chimiques utilisés lors de la préparation des échantillons) ont été retirés de l'analyse.

Les défauts d'hydrolyse par la trypsine ont aussi été pris en compte. La trypsine est une peptidase qui hydrolyse les liaisons peptidiques au niveau d'une lysine ou d'une arginine. Pour tenir compte du fait qu'un certain nombre de ces liaisons n'aient éventuellement pas été rompues lors de

la digestion de nos échantillons, nous avons étendu la recherche aux peptides incluant jusqu'à deux sites Lys-Arg non hydrolysés.

Enfin, nous n'avons considéré que les protéines associées à au moins deux peptides identifiés dans notre échantillon.

L'algorithme Maxquant calcule alors des scores associés à chaque protéine à partir des données brutes et des paramètres spécifiés, ainsi que des p values (Cox 2008). Nous avons utilisé ces dernières pour fixer un taux de fausse découverte (FDR) au niveau des protéines et des peptides à 1 %.

Les protéines conservées ont été quantifiées selon l'algorithme de MaxQuant fournissant une mesure d'intensité normalisée sur l'ensemble des échantillons considérés. Une seconde phase de filtrage consiste ensuite à ne conserver que les protéines présentes dans au minimum 70% des échantillons d'un même groupe.

La matrice de données ainsi obtenue contenait encore de nombreuses données manquantes. Ces dernières ont été imputées à l'aide du logiciel Prostar.

Les analyses statistiques ont ensuite été effectuées en R. Les données dans les conditions patients atteints par la mucoviscidose et contrôles sains ont été comparées par le test t de Student. Des corrections de Benjamini-Hochberg ont été appliquées pour tenir compte des tests multiples. Une p value après correction inférieure à 0,01, et associée à un log2 fold change supérieur à 0,5 a été considérée comme statistiquement significative.

2.3 Résultats

La concentration en protéines se situait entre 0,25 et 0,70 microgrammes par microlitre de sueur, sans différence notable entre les patients atteints de mucoviscidose, les contrôles sains et les hétérozygotes. Après les différentes étapes de filtrages évoquées dans la section précédente, 991 protéines ont été retenues pour l'analyse.

Une première analyse par ACP (figure 23) a permis de mettre en évidence une grande hétérogénéité des patients traités par modulateurs. Alors qu'une partie d'entre eux semblait se rapprocher du profil de patients sains, d'autres sont au contraire apparus plus proches des patients non traités. A l'inverse, les groupes constitués par les contrôles hétérozygotes, les patients non traités et les contrôles sains apparaissaient comme étant bien distincts.

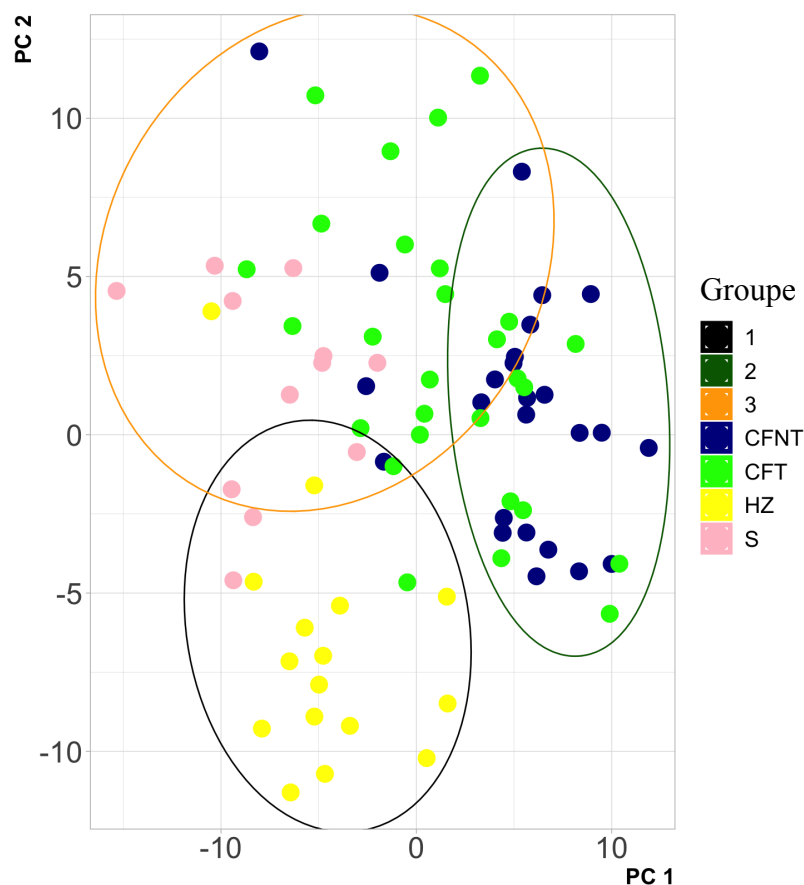


Figure 23 : Composantes 1 et 2 de l'analyse par ACP des données protéomiques des 78 patients de l'étude protéomique. Les ellipses correspondent aux groupes obtenus en utilisant l'algorithme k-means et définissent les groupes 1, 2 et 3. Si les patients HZ, CFNT et S semblent bien se séparer selon la méthode k-means, les patients CFT se répartissent dans les différents groupes. CFNT : patients non traités; CFT : patients traités ; HZ : hétérozygotes ; S : contrôles sains.

La trop grande diversité dans les traitements et dans leur durée rendait difficile analyse et interprétation des données issues des patients traités. Ces derniers ont donc été abandonnés de l'étude.

Les patients hétérozygotes sont apparus très proches des contrôles sains : seules 15 protéines étaient différemment exprimées entre les deux groupes. Pour cette raison, le groupe des hétérozygotes n'a pas non plus été analysé davantage.

L'étude s'est donc concentrée sur les patients non traités et les contrôles sains. 57 protéines sont apparues comme étant différemment exprimées entre ces deux groupes : 25 étaient sur-exprimées dans la sueur des patients porteurs de la mucoviscidose, tandis que 32 étaient à l'inverse sous-exprimées.

Ces 57 protéines sont listées dans l'annexe 8, tandis que la figure 24 représente les résultats de l'analyse différentielle sous la forme d'un graphique en volcan. Une analyse hiérarchique, dont

les résultats sont exposés sur la figure 25, a confirmé que ces protéines séparaient clairement les patients atteints par la mucoviscidose des contrôles sains.

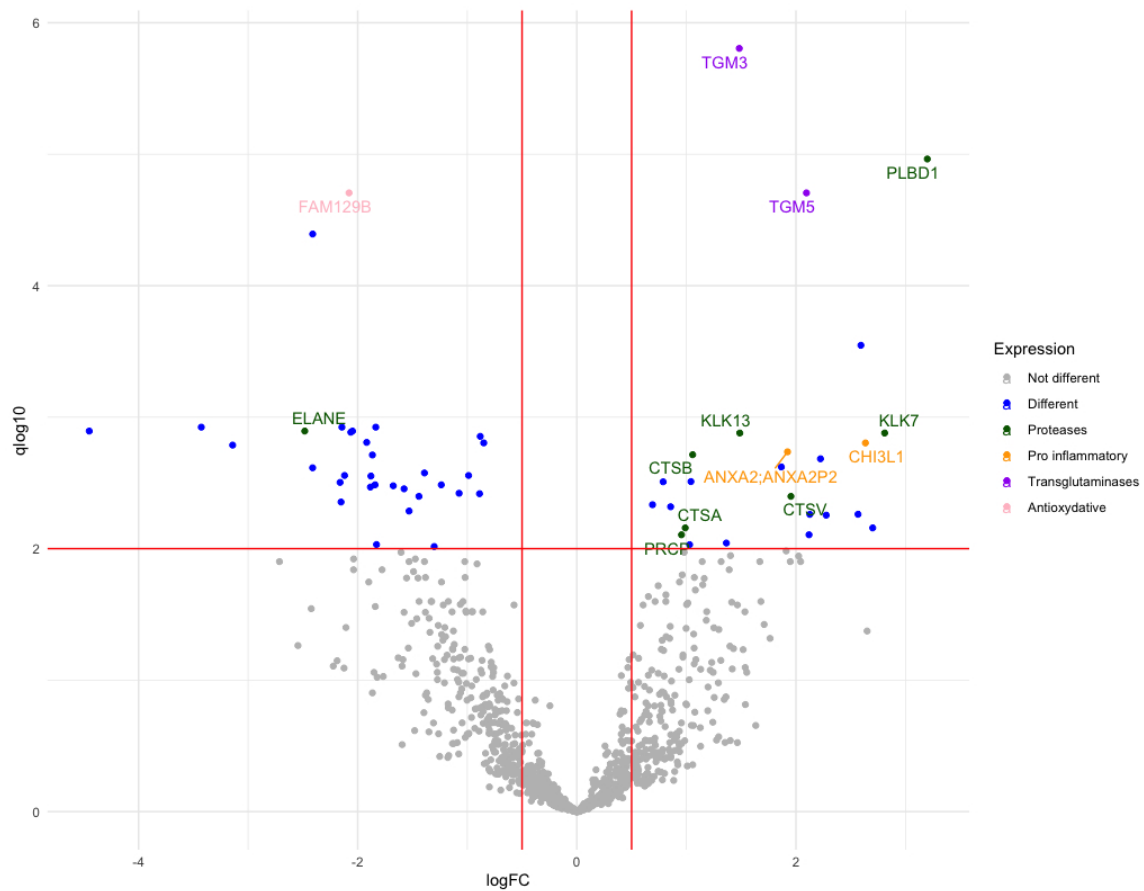


Figure 24 : Graphique en volcan des protéines différentiellement exprimées entre les patients atteints de mucoviscidose (à droite) et les individus sains (à gauche). Des seuils de 0,5 pour le log2 Fold Change (FC) en abscisse, et de 0,01 pour les q values (p values après correction pour tests multiples, exprimées en log10) en ordonnée, sont utilisés pour la significativité et représentés en rouge. Les protéines différentiellement exprimées sont représentées en couleur (protéases : vert ; protéines pro-inflammatoires : orange ; transglutaminases : violet ; protéines antioxydantes : rose ; autres : bleu).

un ppVEMS inférieur à 50%, et/ou la colonisation chronique par *Pseudomonas.aeruginosa* ou *Alcaligenes*, n'ont pas non plus modifié le profil protéomique. Nous avons donc effectué une analyse statistique sur l'ensemble de la cohorte des patients atteints par la mucoviscidose et non traités afin d'accroître la puissance statistique de nos résultats.

Le protéome de la sueur des patients atteints par la mucoviscidose était caractérisé par une augmentation du niveau des (i) protéases (Kallikrein 7 et 13, Phospholipase B domain containing 1, Cathepsin A L2 et B, Lysosomal Pro-X carboxypeptidase, neuraminidase) ; (ii) protéines pro-inflammatoires (Annexin A2, Chitinase-3-like protein 1) ; (iii) cytochrome c ; (iv) Transglutaminases 3 et 5 ; et (v) Toll-interacting protein (TOLLIP). La surpression de cette dernière chez les patients atteints par la mucoviscidose contrastait avec la baisse d'expression de l'Elastase 2, une protéase épidermique, de FAM129B, une protéine antioxydante, et de la Desmoglein-3. Enfin, deux protéines connues pour interagir directement avec CFTR étaient sous-exprimées (Sodium-hydrogen antiporter 3 regulator 1, NHERF1, et Solute Carrier Family 6 Member 14, SLC6A14).

2.4 Discussion de l'étude

Cette étude est à notre connaissance la première à s'intéresser au protéome de la sueur de patients atteints par la mucoviscidose. Nous avons pu réaliser une analyse poussée sur près de 1000 protéines à partir de seulement 20µl de sueur, et ce, grâce à un protocole optimisé impliquant l'acquisition de données par chromatographie en phase liquide suivie d'une analyse par un spectromètre de masse à haute résolution. A notre connaissance, au moment de la rédaction de ce manuscrit, une couverture protéomique d'une profondeur similaire n'a été rapportée que dans deux études précédentes, menées chez des adultes en bonne santé et utilisant des volumes de sueur plus importants. Yu et al ont effectué une analyse sur des échantillons de 10 ml collectés à partir de tissus absorbants placés à différents endroits du corps après un exercice puis combinés (Yu 2017), tandis que Burat et al ont utilisé un volume moyen de 70 microlitres de sueur (Burat 2021). Par ailleurs, aucune de ces deux études ne s'intéressait à la mucoviscidose.

Dans notre cas, nous n'avons utilisé que 20 microlitres de l'échantillon collecté par le test de la sueur. Or, environ 50 microlitres de sueur sont collectés habituellement lors d'un test de la sueur, dont 20 microlitres environ sont utilisés pour l'évaluation de la concentration en ions chlorure dans celle-ci. Nos résultats démontrent ainsi que la quantité de sueur restante peut être utilisée de

manière fiable pour la protéomique de la sueur, y compris chez les nourrissons âgés d'un mois seulement.

Un grand nombre de protéines que nous retrouvons dans notre étude, notamment des protéases ainsi que leurs inhibiteurs respectifs, mais aussi des peptides antimicrobiens et des effecteurs de l'immunité innée de la peau, ont déjà été retrouvées dans la sueur de sujets sains (Csosz 2015, Harshman 2018, Katchman 2018) ou dans des maladies spécifiques comme la schizophrénie ou la tuberculose (Raiszadeh 2012, Adewole 2016). La détection d'autres protéines telles que des protéines du cytosquelette ou des protéines impliquées dans le stress oxydatif, dans la réponse aux protéines non repliées, dans le stress du réticulum endoplasmique, dans le protéasome, ou encore dans les voies du métabolisme, souligne la diversité de la biologie de la sueur (Baechle 2006, Murakami 2002, Park 2011, Schitteck 2001). Bien que certaines études rapportent une corrélation entre la concentration dans le sang et dans la sueur pour certains biomarqueurs spécifiques tels que le glucose (Moyer 2012), la littérature montre des résultats plus mitigés pour d'autres (Baker 2019) et la corrélation entre la concentration dans le sang et dans la sueur n'a pour l'instant été évaluée que pour peu de protéines.

De manière similaire à ce qui avait été montré dans les études précédentes, nous n'avons pas constaté de forte influence du sexe ou de l'âge, même chez les bébés de 1 mois. Nous n'avons pas non plus observé d'influence de la gravité de la maladie en fonction de la génétique ou du profil de la maladie pulmonaire, même si le nombre très limité de patients atteints d'une forme grave de la maladie dans notre cohorte empêche toute conclusion définitive.

Nos résultats apportent un éclairage supplémentaire sur la physiologie de la sueur en montrant un déséquilibre entre l'activité protéolytique et pro-inflammatoire de ce biofluide chez les patients atteints par la mucoviscidose. Un certain nombre de protéases apparaissent surexprimées, notamment les kallikréines 7 et 13. Ces protéases à sérine spécifiques de l'épiderme activent une cascade protéolytique, générant une inflammation cutanée et finalement une altération de la barrière physique de la peau, comme cela a déjà été signalé dans le syndrome de Netherton ou dans le cas de l'eczéma (Kasperek 2017). L'annexine A2 forme un complexe avec le S100A10 impliqué dans une grande variété de processus physiopathologiques, y compris l'activation de l'inflammation induite par les métalloprotéases (Dallacasagrande 2020). D'autres régulateurs de l'immunité innée sont également significativement surexprimés, tels que la Chitinase-3 (CHI3L1) et TOLLIP. La CHI3L1 reflète l'induction des lymphocytes T auxiliaires, l'inflammation induite par l'IL-13 et l'activation des macrophages (Kwak 2019). Il est intéressant de noter que cette protéine est également élevée

dans les expectorations et le sérum des patients souffrants de mucoviscidose et ayant des symptômes pulmonaires avancés (Leonardi 2016).

TOLLIP est un intermédiaire dans la voie de signalisation de l'interleukine (IL)-1 impliquée dans l'inflammation médiée par les récepteurs Toll-like. Cette protéine déclenche la signalisation du facteur nucléaire (NF)- κ B qui est pro inflammatoire (Burns 2000). Toutes ces cascades de signalisation pro-inflammatoires peuvent être associées à un état oxydatif accru, comme le montre l'augmentation de la concentration en cytochrome C, une protéine mitochondriale impliquée dans le transport des électrons et dans la phosphorylation oxydative. Ce phénomène est par ailleurs potentiellement accru par la diminution de la protéine antioxydante FAM129B6 (Cheng 2019).

Nos résultats semblent aussi révéler un défaut de l'intégrité de la barrière cutanée chez les patients atteints par la mucoviscidose. En effet, la sous expression de l'élastase 2 devrait diminuer la quantité de filaggrine clivée, une protéine qui joue un rôle clé dans l'intégrité de la barrière cutanée épidermique (Bonnart 2010, Sandilands 2009). Des études ont ainsi montré que l'absence de filaggrine prédisposait à l'ichtyose vulgaire (la peau devient sèche et rugueuse) et à l'eczéma (Palmer 2006, Weidinger 2006). La sous-expression de la desmoglénine 3, un composant des desmosomes, accentue potentiellement le défaut de barrière cutanée en diminuant la jonction cellule-cellule (Bonné 2003). Enfin, la surexpression des transglutaminases pourrait perturber la formation de la couche externe de l'épiderme en altérant la réticulation des kératines et des protéines structurelles (Rongioletti 2012).

Tous les défauts évoqués précédemment peuvent entraîner une altération de la fonction de barrière cutanée, et ainsi mener à une augmentation de la teneur en eau de la peau lorsque cette dernière y est immergée, ce qui constitue l'une des principales caractéristiques observées dans la KPA (Seidler 2009). En effet, bien que la pathogénie de la KPA ne soit pas entièrement comprise, on suppose qu'elle est liée à des modifications de la teneur en eau de la couche cornée (Rongioletti 2012). Il a déjà été observé que la restauration de la fonction CFTR par VX-770 améliore les symptômes de la KPA, ce qui fournit des preuves cliniques que ce symptôme est lié à un dysfonctionnement de CFTR (Jacobi 2022).

De façon intéressante, dans notre ensemble de données, nous montrons également la sous-expression de 2 protéines connues pour moduler l'expression et la fonction de CFTR à la membrane apicale : NHERF, une protéine connue pour ancrer CFTR à la membrane, et SLC26A14, un transporteur d'acides aminés impliqué dans l'activité du canal F508del-CFTR (Seidler 2009, Ruffin 2020).

Dans l'ensemble, nos résultats, basés sur une évaluation protéomique approfondie, fournissent des preuves d'une activité protéolytique et pro-inflammatoire déséquilibrée de la sueur dans la mucoviscidose, ainsi que d'une altération des voies de signalisation impliquées dans l'intégrité de la barrière cutanée épidermique. Ces observations spécifiques à la mucoviscidose fournissent de nouvelles informations sur la physiopathologie de la KPA et sur la biologie de CFTR. Nous espérons qu'elles pourront être utilisées dans le cadre de la médecine personnalisée pour la mucoviscidose.

Cependant, si notre étude s'est révélée très intéressante dans la compréhension de la physiopathologie de la maladie, en particulier au niveau de la peau, elle s'est en revanche montrée décevante dans la recherche de biomarqueurs dont l'évolution permettrait d'évaluer l'efficacité d'un traitement par les modulateurs de CFTR. Les patients traités par des thérapies constituées d'un ou de deux modulateurs présentaient en effet un profil très variable, et étaient donc difficilement analysables.

2.5 Cas de l'analyse de la trithérapie

Initialement, nous souhaitions aussi étudier l'effet de la trithérapie VX-770/VX-661/VX-445 sur le protéome de la sueur. Pour cela, la sueur de 10 patients supplémentaires, tous éligibles au traitement par la trithérapie, avait été prélevée par deux fois : tout d'abord avant le début du traitement, puis un mois après le début de ce dernier. Néanmoins, chez ces patients, et pour une raison que nous ignorons, le nombre de protéines détectées suite à l'analyse par chromatographie en phase liquide couplée à une spectrométrie de masse était de manière générale bien moins important dans les échantillons prélevés un mois après le début du traitement que dans ceux obtenus pré traitement, comme le souligne la figure 26 ci-après.

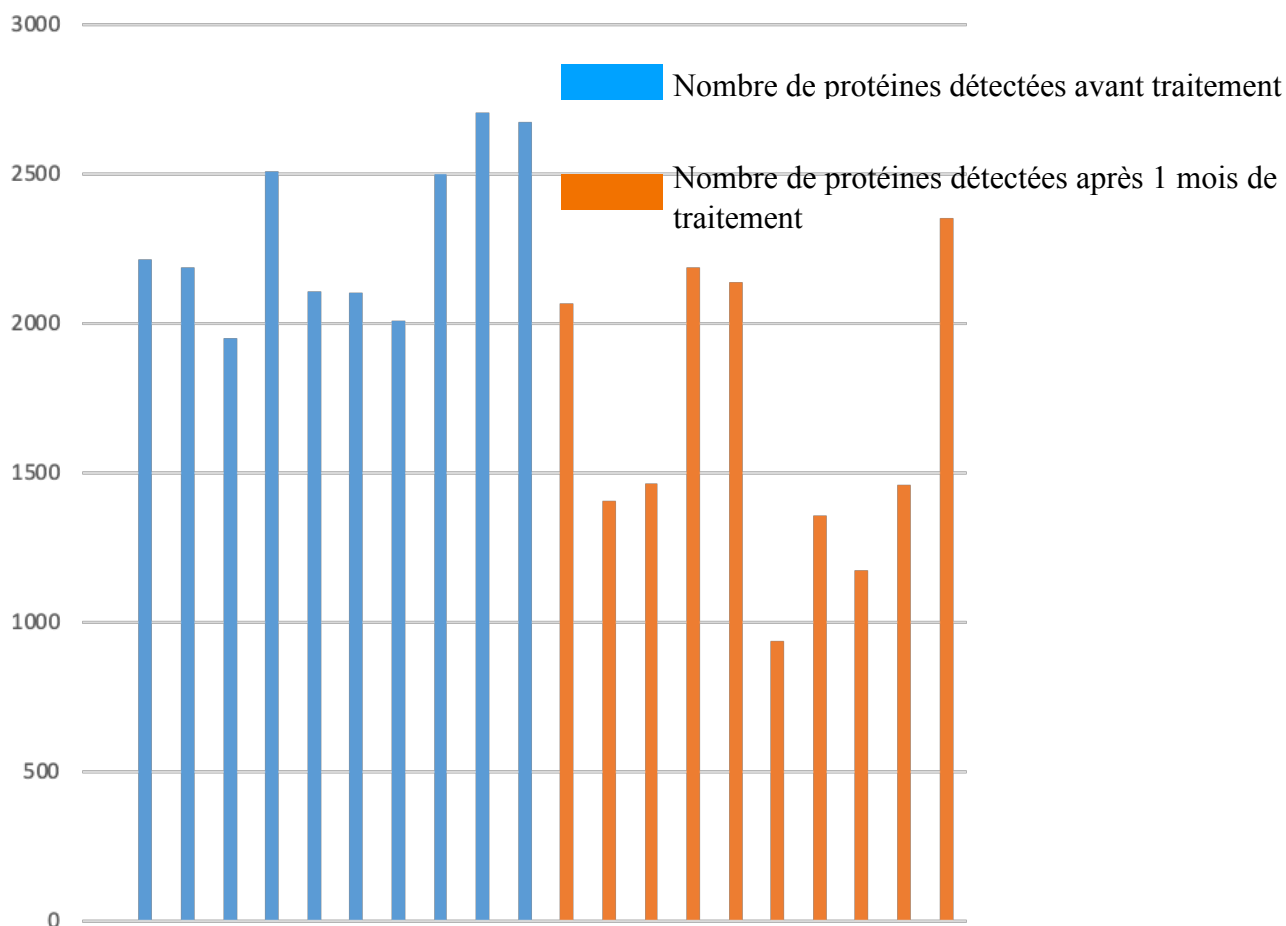


Figure 26 : Nombre de protéines détectées pour les 10 patients traités par la trithérapie. Les barres bleues correspondent aux échantillons prélevés avant traitement, et les barres orange correspondent à ceux obtenus après un mois de traitement. L'ordre des patients est le même pour les deux couleurs de barre (la première barre bleue et la première barre orange correspondent au même patient)

Plusieurs hypothèses ont été envisagées pour expliquer ce phénomène. Nous avons notamment vérifié que la quantité de sueur collectée chez ces patients était sensiblement la même dans les deux prélèvements effectués, ce qui était effectivement le cas. La plus faible quantité de protéines détectées ne pouvait donc pas s'expliquer par une quantité de matériel initial plus faible.

Malgré ces limitations, nous avons essayé d'utiliser ces données dans l'espoir de trouver des effets caractéristiques de la trithérapie. Le jeu de données étudié comportait de nombreuses données manquantes, et il était donc nécessaire d'être très prudent lors de l'imputation de ces dernières.

Les jeux de données en protéomique sont très sujets au problème des données manquantes, c'est-à-dire associées à des intensités très faibles. Certaines de ces données manquantes correspondent à des protéines non détectées techniquement mais qui sont en fait exprimées à des niveaux significatifs dans les échantillons, que l'on peut appeler des « zéros techniques ». D'autres données manquantes correspondent à des protéines qui sont effectivement exprimées à de très faibles niveaux dans les échantillons, et que l'on peut appeler de « zéros biologiques ». Plusieurs

outils ont été développés dans le but de distinguer ces deux situations. Ainsi, les données de protéomique ne peuvent pas être considérées comme étant "Missing à random" (MAR) ni "Missing Completely at Random" (MCAR), mais sont au contraire le plus souvent "Missing Not At Random" (MNAR). Cela signifie que la probabilité qu'une donnée soit manquante n'est pas la même pour tous les gènes et tous les échantillons. L'enjeu des logiciels d'imputation de données manquantes en protéomique est donc de déterminer, parmi les données manquantes, celles qui sont des zéros techniques, et celles qui sont des zéros biologiques.

Plusieurs études ont analysé l'impact du choix de la méthode d'imputation sur les résultats obtenus lors d'une analyse différentielle en protéomique (Liu 2020, Jin 2021). Liu et al. concluent que le choix de la méthode d'imputation n'influe que faiblement sur les résultats si la préparation a été correctement effectuée, et montrent notamment qu'une simple imputation systématique des données manquantes par des valeurs faibles (inférieures au minimum des données présentes) donne des résultats satisfaisants. Bien que ce choix de méthode d'imputation ne fasse pas l'unanimité, Jin et al. recommandent à l'inverse d'utiliser une méthode essayant de séparer les données MAR/MCAR des données MNAR, le fait que la méthode d'imputation ait un impact limité sur les résultats n'est pas remis en cause.

Dans le cadre de notre étude avec la trithérapie, nous avons souhaité vérifier l'effet de cette imputation sur l'interprétation des résultats. Nous avons utilisé deux méthodes distinctes proposées par deux logiciels, Prostar (Wieczorek 2017) et Perseus (Tyanova 2018).

L'imputation proposée par Prostar se rapproche des recommandations de Jin et al. Par ailleurs, c'est cette méthode d'imputation que nous avons choisi d'utiliser dans la première partie de notre étude. Elle consiste à rechercher quelles données sont MAR, et les quelles sont MNAR, avant de réaliser deux imputations différentes.

Les échantillons sont tout d'abord séparés en groupes, selon l'analyse différentielle qui doit être effectuée. Dans notre cas, la séparation a été réalisée entre les échantillons prélevés avant traitement et ceux obtenus 1 mois après le début de celui-ci. Puis, pour chaque protéine détectée dans au moins un des échantillons (et après les différentes règles de filtrages expliquées dans la première partie de cette étude), on vérifie si elle est totalement manquante dans un des groupes. On considère alors dans un tel cas qu'il s'agit de données MNAR : si une protéine n'est jamais détectée dans un groupe, cela signifie qu'elle n'était effectivement présente qu'en quantité minimale dans les échantillons qui le composent. A l'inverse, si la protéine est détectée dans au moins l'un des échantillons de chaque groupe, alors on considère les données manquantes pour cette protéine

comme MAR : si une protéine a été détectée dans certains des échantillons d'un groupe, alors elle aurait dû l'être pour tous les échantillons de ce même groupe.

Les données MNAR sont ensuite simplement imputées avec la valeur minimale présente dans le jeu de données. Les données MAR sont quant à elles imputées selon un modèle linéaire, à partir des protéines dont le profil d'expression est le plus proche de celui de la protéine pour laquelle l'imputation est réalisée.

L'imputation proposée par Perseus est beaucoup plus simple et se rapproche des recommandations de Jin et al.. Toute valeur manquante est considérée comme MAR et est imputée avec une valeur tirée à partir d'une Gaussienne dont la moyenne est inférieure au minimum des expressions détectées pour la protéine dont on réalise l'imputation de données.

Deux problèmes majeurs se sont posés dans le cadre de notre étude sur les patients traités par la trithérapie. Nous avons décidé dans un premier temps d'utiliser la même méthode d'analyse que dans lors de notre étude comparative entre les patients non traités atteints par la mucoviscidose et les contrôles sains. Nous réalisons un test t de Student (apparié ici, puisqu'il s'agit des mêmes patients avant et après début du traitement), puis nous prenions en compte l'aspect tests multiples en réalisant une correction de Benjamini Hochberg, nous donnant une q value. Cependant, après cette correction, seules 7 protéines présentaient une q value inférieure à 0.05 dans le cas d'une imputation des valeurs manquantes via Prostar, et aucune dans le cas d'une imputation par Perseus. Il était impossible de tirer des conclusions de cette analyse dans de telles conditions.

Il a donc été décidé de changer la méthode de correction pour tests multiples, et de privilégier une méthode proposée par le logiciel Perseus. De cette manière il a été possible de relever plus de 30 protéines différentiellement exprimées entre les patients pré et post début du traitement. La méthode proposée par Perseus se base sur des réassignations multiples aléatoires des groupes associés à chaque échantillon de l'étude. Le nombre de gènes ainsi trouvés « par chance » comme étant significatifs est comparé au nombre de gènes trouvés dans l'étude initiale et utilisé comme métrique d'ajustement des p values.

Néanmoins, nous nous sommes retrouvés confrontés à un second problème. Il est apparu que le choix de la méthode d'imputation influait grandement sur les résultats obtenus lors de l'analyse différentielle, comme le montre la figure suivante (figure 27) :

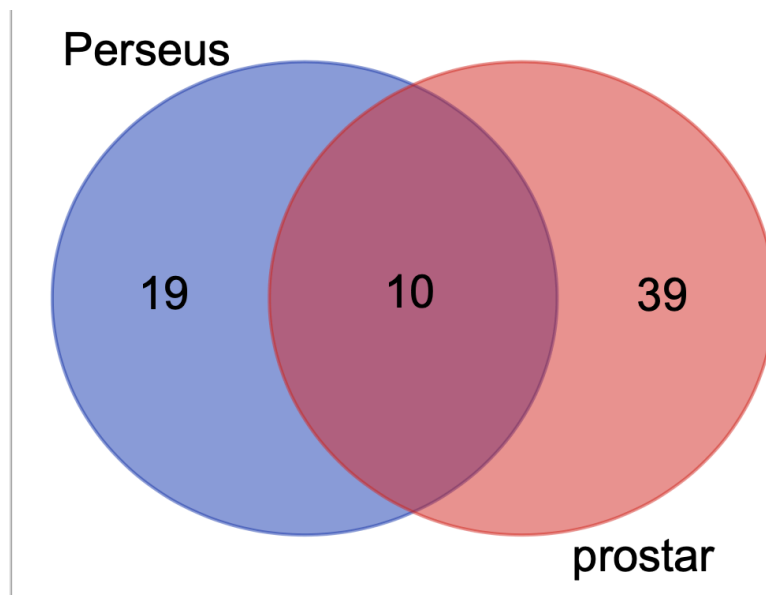


Figure 27 : nombre de protéines détectées lors d’une analyse différentielle à partir des échantillons de 10 patients avant et après le début du traitement par la trithérapie, en fonction de la méthode d’imputation utilisée.

Seules un tiers des protéines détectées comme étant différentiellement exprimées en utilisant la méthode d’imputation de Perseus l’étaient aussi en utilisant la méthode d’imputation Prostar. Cela ne représentait par ailleurs qu’un cinquième des protéines relevées en utilisant cette méthode.

Les 10 protéines retrouvées dans les deux cas avaient peu de lien entre elles, et étaient en outre très difficilement interprétables. Ainsi, parmi les protéines les plus différentiellement exprimées, on retrouvait SET, une chaperone d’histones, la nucleoline, impliquée dans la synthèse et la maturation des ribosomes, ou encore FUCA1, impliquée dans la dégradation de glycoprotéines et de glycolipides contenant du fucose.

Comme on peut le constater sur la figure 26, pour 4 des patients la différence de protéines détectées entre l’échantillon pré traitement et celui post traitement était faible. Nous avons donc essayé de reprendre les mêmes analyses à partir de ces 4 patients uniquement. Malheureusement, le trop faible nombre d’individus considérés a mené à l’impossibilité de détecter la moindre protéine différentiellement exprimée.

Les échantillons ont alors été repassés une seconde fois dans l’ensemble chromatographie en phase liquide suivi du spectromètre de masse. Malheureusement les données obtenues sont restées très similaires à celle de la première série d’analyse, à la fois en termes d’expression et de nombre de protéines détectées.

L'analyse de la protéomique de patients traités par la trithérapie est donc restée sans conclusions claires.

2.6 Conclusion

Cette étude a permis de mettre en avant des pistes sur l'origine de la KPA chez les patients atteints par la mucoviscidose, et en ce sens elle est importante pour mieux comprendre la maladie. En revanche, les limitations techniques ne nous ont pas permis de pousser notre étude dans le cas de patients traités, que ce soit par des monothérapies telles que VX-770, des bithérapies telles que VX-770/VX-809, ou la trithérapie VX-770/VX-661/VX-445. Pour ce dernier cas, des analyses plus poussées sont encore nécessaires afin de déterminer précisément les causes de ces limitations. Il est par ailleurs possible que la plus petite quantité de protéines détectées chez les patients traités par la trithérapie, qui a mené aux difficultés techniques rencontrées, soit en réalité associée à un effet particulier du traitement.

Si l'analyse protéomique de la sueur s'est révélée être un relatif échec dans notre recherche de biomarqueurs pour évaluer l'effet de la trithérapie, nous n'avons pas abandonné notre recherche en ce sens et avons décidé de concentrer nos efforts sur l'appareil respiratoire, qui contient les principaux organes affectés dans la mucoviscidose. Nous avons ainsi repris notre logique d'analyse différentielle en protéomique entre des contrôles sains, des patients non traités et des patients traités, que nous avons cette fois-ci appliqué au cas de cellules provenant du nez des patients. Ce fut l'objet de l'étude décrite dans la partie suivante.

3. Analyse multi-omique de cellules primaire de patients, évaluation des effets de traitements par modulateurs

Les résultats associés à cette sections n'ont pas encore été publiés au moment de la rédaction de ce manuscrit.

3.1 Introduction :

L'étude précédente a mis en évidence l'intérêt d'utiliser des approches omiques dans la mucoviscidose afin de mieux comprendre la physiopathologie de cette maladie. Elle n'a malheureusement pas permis d'apporter des informations supplémentaires sur les traitements par modulateurs, mais cela était essentiellement dû à des limitations techniques du fait de l'analyse d'un biofluide, la sueur. En pratique, les études omiques sont le plus souvent réalisées à partir de matériel génétique provenant de cellules de patients.

Dans le cadre de la mucoviscidose, la maladie est associée à une bronchopathie muco-obstructive ainsi qu'à des infections et à des inflammations chroniques, conduisant progressivement à des lésions pulmonaires et à un déclin de la fonction respiratoire. Ainsi, l'étude précise du contenu des cellules épithéliales des voies respiratoires apparaît comme particulièrement intéressante pour la mucoviscidose.

Les récentes avancées dans la thérapie de la mucoviscidose avec une combinaison de 3 modulateurs de CFTR (VX-770/VX-661/VX-445) ont considérablement amélioré le pronostic des patients porteurs de la mutation F508del. Cependant, le mécanisme d'action exact de la trithérapie reste incomplètement compris. La comparaison en protéomique et en transcriptomique de cellules de patients traités avec celles de patients non traités par ces modulateurs pourrait améliorer notre compréhension de leur mécanisme d'action, et potentiellement mettre en évidence des protéines ciblées différentes de CFTR, dont le ciblage spécifique par des molécules optimisées pour ces protéines pourrait s'avérer bénéfique pour les patients.

De plus, si la trithérapie est une avancée significative pour les patients porteurs de mutations de classe II ou supérieure, ses effets sont beaucoup plus limités chez les patients porteurs de mutations de classes I. Cependant, aucune étude n'a à ce jour étudié l'effet de la trithérapie chez de tels patients via une approche omique. Une telle analyse pourrait mettre en évidence un potentiel effet des modulateurs indépendant de CFTR, ouvrant la voie à la recherche de nouveaux traitements

ouverts aux patients porteurs de mutations de classe I, et pouvant par ailleurs élargir l'arsenal thérapeutique disponible pour les patients porteurs de mutations de classe II ou supérieure.

Dans la présente étude, nous avons dans un premier temps analysé l'effet de la combinaison VX-770/VX-661/VX-445 au niveau transcriptomique et protéomique sur les cellules épithéliales nasales humaines primaires (HNE) de 10 enfants atteints par la mucoviscidose et homozygotes pour la mutation F508del. Dans un second temps nous avons réalisé la même étude à partir de cellules HNE provenant de 5 patients homozygotes pour la mutation G542X exprimant donc une protéine tronquée, et 4 autres patients porteurs de mutations de classe I. Les cellules HNE ont été différenciées en un épithélium pseudo-stratifié dans des conditions d'interface air-liquide sur des filtres microporeux et traitées pendant 48 heures avec du VX-770/VX-661/VX-445 ou un véhicule (DMSO).

Bien que les échantillons aient montré une grande variabilité inter-patients globale dans les données transcriptomiques et protéomiques, témoignant essentiellement du fait que les fonds génétiques de ces patients sont différents, le traitement avec le VX-770/VX-661/VX-445 a induit une expression différentielle de 37 protéines et 268 gènes chez les patients homozygotes pour la mutation F508del. Bien qu'il y ait eu peu de chevauchement en termes de gènes individuels et de protéines, l'analyse des voies dans les études transcriptomiques et protéomiques a montré des résultats cohérents.

De manière inattendue, les principales voies régulées à la hausse par le VX-770/VX-661/VX-445 impliquaient les fonctions mitochondriales : cycle de l'acide tricarboxylique, bêta-oxydation des acides gras et phosphorylation oxydative. Ces résultats suggèrent une modification du métabolisme mitochondrial induite par le traitement et qui pourrait être bénéfique aux patients atteints par la mucoviscidose. De manière encore plus intéressante, la même tendance est apparue chez les patients porteurs de mutations de classe I, suggérant que ce mécanisme ciblé par les modulateurs et potentiellement bénéfique pour les patients est indépendant de CFTR.

3.2 Matériels et méthodes

3.2.1 PATIENTS DE L'ETUDE

L'étude s'est décomposée en deux parties. Dans un premier temps, nous avons pris en considération les cellules provenant de 10 patients homozygotes pour la mutation F508del et réalisé

des analyses en transcriptomique et en protéomique. Afin de limiter au maximum les variations qui ne seraient pas liées aux traitements, tous les patients avaient des âges similaires au moment du traitement (entre 8 et 10 ans), et l'équilibre homme/femme était respecté avec 5 garçons et 5 filles.

Dans un second temps, nous nous sommes intéressés à des échantillons provenant de 9 patients porteurs de mutation de classe I, dont 5 étaient homozygotes pour la mutation G542X, la plus courante des mutations de cette classe, et nous avons réalisé une analyse protéomique uniquement. Nous avons traités ou non les cellules avec la trithérapie VX-770/VX-661/VX-445 afin d'évaluer l'effet de ces traitements sur des épithéliums de patients. Tous les individus considérés étaient âgés entre 8 et 12 ans, et 5 des 9 patients étaient des garçons.

3.2.2 CULTURE CELLULAIRE

Si les cellules épithéliales bronchiques sont les plus représentatives de la maladie, leur difficulté d'accès les rend difficilement utilisables en pratique: il est nécessaire d'extraire une portion du poumon des patients. Notre choix s'est donc porté sur l'analyse de cellules épithéliales nasales, beaucoup plus faciles d'accès et dont le prélèvement n'est associé qu'à une gêne très légère pour le patient. De plus, le fonctionnement des cellules épithéliales nasales est proche de celui des cellules épithéliales bronchiques, et peut donc bien refléter les dysfonctionnements présents dans les poumons de patients atteints par la mucoviscidose.

Brossage nasal

Les cellules de patients sont obtenues par brossage nasal. Celui-ci est effectué en s'assurant au préalable que le patient ne présente pas d'infection aiguë.

Après s'être mouché, la muqueuse des deux narines du patient est anesthésiée à l'aide de mailles de coton imbibées de xylocaïne. La paroi médiane des deux narines est ensuite brossée délicatement à l'aide d'une brosse à cytologie. Enfin, la brosse est trempée dans un tube contenant un milieu de rinçage, et agitée doucement pour détacher les cellules. Le même processus est répété dans la deuxième narine.

Une fois toutes les cellules détachées, des antibiotiques (Pénicilline, Streptomycine, Tazocilline, Amphotéricine B et Colimycine) sont ajoutés au milieu de rinçage afin de limiter au maximum le développement de pathogènes qui auraient pu être prélevés avec les cellules. En cas de présence d'une infection pulmonaire chez certains des patients, la quantité d'antibiotiques utilisés pour ces derniers est augmentée.

Isolement des HNE

Les cellules dans le milieu de rinçage sont placées dans un tube pour être centrifugées. Elles sont ensuite récupérées et remises en suspension dans de la trypsine afin de les dissocier. Après une dizaine de minutes (temps nécessaire à la dissociation des cellules par la trypsine), la réaction est arrêtée par un ajout de milieu de culture, puis les cellules sont à nouveau centrifugées pour retirer la trypsine.

Elles sont ensuite re-suspendues dans un milieu d'amplification etensemencées dans une fiole de 25 cm² recouverte de collagène (permettant aux cellules d'adhérer sur la flasque). Cette dernière est ensuite placée dans un incubateur à 37 °C et 5 % de CO₂ afin de reproduire les conditions physiologiques du corps humain.

Le milieu est remplacé chaque jour jusqu'à ce que les cellules atteignent 80 à 90 % de confluence (i.e remplissent 80 à 90% de la flasque), ce qui prend en général entre 3 et 10 jours, en fonction de la densité initiale de cellules.

Amplification et passage des cellules HNE

Lorsque les cellules atteignent 80 à 90% de confluence, le milieu de culture est retiré et les cellules sont lavées avec du PBS (afin de s'assurer que la trypsine puisse faire effet par la suite). Puis de la trypsine est ajoutée dans la flasque et laissée pendant une dizaine de minutes afin de dissocier les cellules. Ces dernières sont ensuite décollées de la surface de la flasque en tapotant fermement sur celle-ci.

La réaction enzymatique est arrêtée par ajout de milieu de culture, puis la trypsine est retirée par centrifugation. Les cellules sont alors re-suspendues dans du milieu de culture et placées dans plusieurs flasques de taille supérieure (dont le nombre dépend de la quantité de cellules obtenues) et recouvertes de collagène.

Cryoconservation des cellules épithéliales nasales primaires

Le même processus de culture est répété dans les nouvelles flasque en changeant le milieu de culture tous les deux jours. Une fois que les cellules ont atteint 80 à 90% de confluence, elles sont à nouveau extraites des flasques avec la méthode décrite précédemment.

Les cellules sont alors comptées, puis re-suspendues dans un milieu de congélation, avant d'être congelées en les amenant progressivement à une température de -80°C.

Décongélation des cellules congelées

Les cellules sont retirées de leur lieu de stockage et rapidement amenées à 37°C en les plongeant dans un bain-marie. Une fois à bonne température, le milieu de congélation est retiré par centrifugation et remplacé par du milieu d'amplification, puis les cellules sont à nouveau placées dans des flasques afin de reprendre leur développement. Une nouvelle phase d'expansion cellulaire s'effectue alors, de manière similaire à ce qui a été décrit précédemment.

Différenciation des cellules épithéliales nasales humaines à l'interface air-liquide

Les cellules cultivées en flasques ne sont pas différenciées. Or nous souhaitons réaliser une analyse à partir de cellules qui soient les plus proches possibles de celles présentes au niveau d'un épithélium respiratoire dans le corps humain. Une phase de différenciation est donc nécessaire afin de recréer un épithélium en culture.

Cette différenciation est réalisée sur de petits filtres poreux (surface de 0,33 cm²). Dans un premier temps, 330 000 cellules sontensemencées sur chaque filtre en présence de milieu de culture à la fois au niveau apical et basal.

Le milieu de culture reste présent pendant trois jours au niveau apical, empêchant la différenciation des cellules mais permettant à ces dernières d'adhérer au filtre et de continuer à se développer jusqu'à le recouvrir en totalité.

Après trois jours, le milieu apical est aspiré, et les cellules sont cultivées à l'interface air-liquide (présence de milieu en basal, mais d'air en apical) afin d'imiter le comportement d'un véritable épithélium.

Les cellules sont alors différenciées pendant trois semaines en renouvelant le milieu basal tous les deux à trois jours.

Traitements par modulateurs de CFTR

A partir de 48 heures avant l'extraction des cellules pour analyse, le milieu de culture basal est modifié afin qu'il contienne des modulateurs de CFTR. Dans notre étude, nous nous sommes concentrés sur l'effet de la trithérapie VX-445/VX-661/VX-809. Afin d'imiter des conditions réelles de traitement chez le patient, nous avons utilisé VX-445 et VX-661 à une concentration finale de 3 µM, et VX-770 à 100 nM, que nous avons dissouts dans du DMSO. Le milieu est renouvelé après 48h.

Afin d'obtenir des échantillons contrôles, un second milieu basal contenant la même quantité de DMSO que dans le milieu correcteur, mais sans les modulateurs, a été préparé. Celui-ci a été utilisé pour la moitié des filtres, devenant ainsi les échantillons contrôles.

Le processus complet de culture est résumé dans la figure 28.

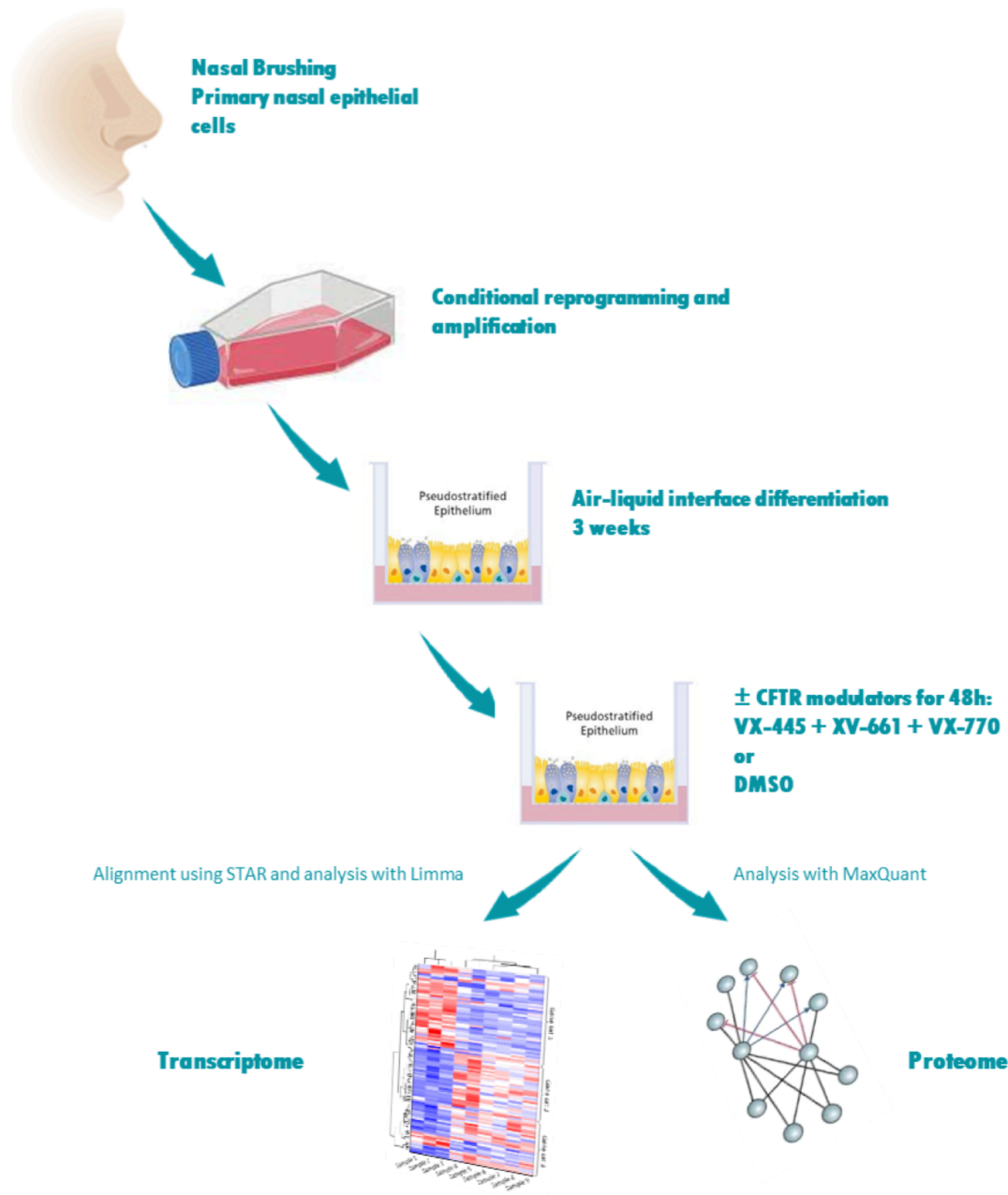


Figure 28 : Résumé du processus de culture cellulaire. Des cellules primaires sont récoltées au cours d'un frottis nasal puis mises en cultures sur flasques. Elles sont ensuite différenciées en un épithélium et traitées ou non par la trithérapie, puis une analyse de leur protéome et de leur transcriptome est réalisée.

Extraction de l'ARN

Pour chaque patient, 12 filtres étaient consacrés à l'analyse ARN : 6 contenaient des cellules traitées par modulateurs et 6 des cellules traitées uniquement au DMSO.

Dans un premier temps, les cellules de chaque filtre sont lysées et homogénéisées dans 1mL de réactif TRIzol™, une solution chimique séparant l'ARN de l'ADN et des protéines.

0,2 mL de chloroforme pour 1 mL de TRIzol™ sont ajoutés, puis mélangés soigneusement en agitant les tubes, puis une centrifugation a lieu, menant à l'apparition de 3 phases : une phase inférieure rouge de phénol-chloroforme, organique, qui contient les protéines, une interphase contenant l'ADN, et une phase supérieure aqueuse incolore contenant l'ARN. La phase supérieure est alors pipetée et transférée dans un nouveau tube.

Les ARN totaux sont par la suite isolés à l'aide du kit RNeasy (QIAGEN) comprenant une étape de traitement par la DNase (afin de s'assurer qu'aucun brin d'ADN n'a été extrait par erreur). La qualité de l'ARN est évaluée par électrophorèse capillaire en utilisant les réactifs High Sensitivity RNA avec le Fragment Analyzer (Agilent Technologies) et la concentration de l'ARN est mesurée en utilisant à la fois la spectrophométrie avec le Xpose (Trinean) et l'électrophorèse capillaire du Fragment Analyzer. Tous les échantillons ont présenté une très bonne pureté ainsi que des concentrations suffisantes pour analyse (au delà de 20 nanogrammes d'ARN par microlitre de solution). Les scores RIN étaient en effet tous supérieurs à 9, alors qu'ils sont considérés comme acceptables à partir de 8.

3.2.3 ANALYSE RNA-SEQ

Les bibliothèques RNA-seq ont été préparées à partir de 500 ng d'ARN total en utilisant le kit Universal Plus mRNA-Seq (Nugen-Tecan) comme recommandé par le fabricant. Avec ce kit, des bibliothèques RNA-seq à brins peuvent être préparées à partir de 10 ng à 1µg d'ARN total.

Une description détaillée de ce processus est donnée en introduction. En bref, les ARNm ont été capturés avec des billes magnétiques polyA+ à partir de l'ARN total. Les ARNm ont été fragmentés chimiquement. Un ADNc simple brin et un second brin ont été produits, puis ligaturés à des adaptateurs compatibles. Les ADNc produits ont été amplifiés par PCR après un test initial visant à évaluer le nombre approprié de cycles de PCR à appliquer à chaque échantillon.

Les bibliothèques RNA-seq ainsi obtenues ont été séquencées sur un NovaSeq6000 d'Illumina. Un total d'environ 50 millions de paires de séquences a été produit par bibliothèques.

Etant donné que l'étude des patients homozygotes pour la mutation F508del comportait 10 individus, soit 60 échantillons (3 réplicats pour les conditions traitées et non traitées), la fabrication des banques du projet a été coupée en 2 séries de banques, car seuls 48 échantillons pouvaient être traités en parallèle au maximum. Cette séparation en deux séries pouvait entraîner des effets de batch dont il a fallu tenir compte lors de l'analyse informatique. Plusieurs contrôles ont en outre été réalisés sur les deux séries afin de réduire au maximum les disparités de quantités de données produites entre échantillons.

Pour chaque librairie RNA-seq générée, la qualité du séquençage a d'abord été évaluée avec FASTQC. Les fichiers FASTQ ont ensuite été alignés sur le génome de référence HG38 (Genome Reference Consortium Human Build 38) en utilisant l'aligneur STAR (Dobin 2013). STAR a également été utilisé pour générer la matrice de comptage, en comptant les lectures par gène et par échantillon.

Nous avons utilisé Limma (Ritchie 2015) pour effectuer l'analyse différentielle. Les gènes faiblement exprimés ont d'abord été filtrés à l'aide de la fonction `filterByExpr` de la librairie R `edgeR` (Robinson 2010). En bref, le filtrage conserve les gènes dont le nombre par million est supérieur à 10 dans au moins la moitié des échantillons (c'est-à-dire au moins la taille d'un groupe). Les données ont ensuite été normalisées (Robinson 2010) en utilisant la fonction `calcNormFactors` de `edgeR`. L'hétéroscédasticité a été supprimée des données à l'aide de la fonction `Voom` de Limma. Enfin, un modèle linéaire a été ajusté pour comparer les échantillons traités et non traités tout en tenant compte des échantillons appariés (c'est-à-dire des échantillons provenant d'un même patient), des réplicats techniques et de l'effet de batch.

3.2.4 ANALYSE PROTEOMIQUE

Pour la protéomique, les données ont été analysées avec MaxQuant version 2.0.1.0 et recherchées avec le moteur de recherche Andromeda dans la base de données UniProtKB Homo sapiens (version 02-2021, 20396 entrées). La déviation de masse a été fixée à 3 ppm, et deux sites de clivage manqués ont été autorisés. La carbamidométhylation (Cys) a été définie comme une modification fixe, tandis que l'oxydation (Met) et l'acétylation N-terminale ont été définies comme des modifications variables. Les taux de fausse découverte au niveau des protéines et des peptides ont été fixés à 1%. Les protéines ont été quantifiées selon l'algorithme MaxQuant en utilisant des mesures d'intensités; la quantification des protéines a été obtenue en utilisant au moins 2 peptides par protéine. Pour plus de détail sur le processus d'analyse en spectrométrie de masse et

chromatographie en phase liquide, se référer à la partie précédente et à la section introductive de ce manuscrit.

Dans le cas des patients homozygotes pour la mutation F508del, le protocole d'extraction ayant été différent pour les deux premiers échantillons, un nombre moins important de protéines ont été détectées dans ceux-ci par rapport au reste des échantillons analysés. Par conséquent, ces deux échantillons ont été retirés de l'étude protéomique. Le filtrage et l'imputation des données ont été effectués à l'aide du logiciel Prostar (Wieczorek 2017). Les données ont été imputées à l'aide de l'algorithme des moindres carrés structurés pour les valeurs partiellement observées et de l'algorithme DetQuantile pour les données manquantes dans une condition entière (de manière similaire à ce qui a été décrit dans la section précédente). Les protéines ont été retenues si elles étaient détectées dans au moins 70 % des échantillons d'un groupe de l'analyse (c'est-à-dire 5 échantillons sur 8 dans un groupe dans le cas des patients porteurs de la mutation F508del). Les analyses statistiques ont été effectuées avec R. Les données ont été comparées par le test de Student pairé lorsque nécessaire.

3.2.5 ANALYSE INFORMATIQUE

Dans les études transcriptomiques et protéomiques, des corrections de Benjamini-Hochberg ont été appliquées pour tenir compte des tests multiples. Pour l'analyse de l'expression différentielle des gènes, une valeur $q < 0,05$ associée à une variation absolue du fold change $> 1,1$ a été considérée comme statistiquement significative. Pour l'analyse des voies, une valeur $q < 0,05$ a été considérée comme significative en transcriptomique, et une valeur $q < 0,01$ a été considérée comme significative en protéomique.

L'analyse des voies a été réalisée à l'aide de la librairie R FGSEA (Korotkevich 2016). Plus de 1000 voies ont été testées, y compris les bases de données de voies Hallmark (Liberzon 2015), PID (Schaefer 2009), Ingenuity Pathways Analysis (Krämer 2014) et Dorothea (Garcia-Alonso 2019).

FGSEA est une version accélérée de l'algorithme GSEA (Gene Set Enrichment Analysis). Ce dernier permet, comme son nom l'indique, d'analyser l'expression de groupements de gènes. Dans une analyse donnée, les gènes sont triés par ordre croissant de p values, et le rang moyen des gènes d'un même groupement est utilisé afin de calculer une p value à l'échelle du groupement.

Afin de compléter l'analyse par approche GSEA, une seconde analyse de voies a été réalisée à l'aide de l'algorithme rRoma, dont les détails sont exposés dans la section suivante. Pour rRoma,

les analyses ont été menées à partir de la base de données Hallmark, détaillée dans la section suivante.

3.3 Résultats

A partir des 8 individus conservés pour le cas des patients homozygotes pour la mutation F508del, 5742 protéines ont été considérées dans l'analyse protéomique. 37 protéines ont été trouvées différentiellement exprimées entre les échantillons VX-770/VX-661/VX-445 et DMSO (Tableau 9 en annexe), dont 34 étaient plus abondantes dans le cas VX-770/VX-661/VX-445. En termes de voies, 34 voies ont été trouvées différentiellement exprimées, dont 17 étaient plus abondantes dans le cas VX-770/VX-661/VX-445. Les voies les plus significatives comprenaient les voies liées aux mitochondries et au cycle de Krebs.

16701 gènes ont été considérés dans l'analyse transcriptomique. 208 ont été trouvés différentiellement exprimés. 136 étaient plus abondants dans VX-770/VX-661/VX-445, tandis que 72 étaient plus abondants dans le cas contrôle avec le DMSO. En termes de voies, 52 voies ont été trouvées différentiellement exprimées, dont 24 étaient régulées à la hausse avec VX-770/VX-661/VX-445. Les voies comprenaient la réponse à l'interféron Alpha et à l'interféron Gamma, Stat1 et Stat2, la phosphorylation oxydative, le métabolisme des acides gras, le dysfonctionnement mitochondrial et le cycle de Krebs. De manière surprenante, les voies de l'interféron ont été régulées à la hausse par le VX-770/VX-661/VX-445.

Seuls 2 gènes ont été trouvés différentiellement exprimés dans les deux études (CPT1A, EPS8L1). Cependant, l'analyse par voies avec GSEA s'est révélée être beaucoup plus intéressante, de nombreuses voies étaient partagées par les deux études, notamment la phosphorylation oxydative, le métabolisme des acides gras et le dysfonctionnement mitochondrial (Figure 29). L'analyse avec rRoma a confirmé ces premiers résultats, mettant aussi en avant des voies mitochondriales (résultats détaillés en annexe 10).

Transcriptomic

PATHWAY	p value	Effect of CFTR modulators
OXIDATIVE PHOSPHORYLATION	4,73E-25	upregulation
INTERFERON GAMMA RESPONSE	3,57E-17	upregulation
INTERFERON ALPHA RESPONSE	1,03E-13	upregulation
STAT2	2,49E-09	upregulation
HYPOXIA	3,41E-07	downregulation
ADIPOGENESIS	4,68E-06	upregulation
UNFOLDED PROTEIN RESPONSE	5,80E-05	downregulation
TNFA SIGNALING	9,62E-05	downregulation
FATTY ACID METABOLISM	5,01E-04	upregulation
STAT1	1,86E-03	upregulation
SPI1	3,03E-03	upregulation
PPARG	3,78E-03	upregulation
ARF PATHWAY	4,28E-03	downregulation
TCA CYCLE	6,76E-03	upregulation
Graft-versus-Host Disease Signaling	6,90E-03	upregulation
P53 PATHWAY	7,04E-03	downregulation
XENOBIOTIC METABOLISM	7,72E-03	upregulation
ERBB1 PATHWAY	1,11E-02	downregulation
LPS/IL-1 Mediated Inhibition of RXR Function	1,25E-02	upregulation
BACH1	1,26E-02	upregulation
NCOA3	1,36E-02	upregulation
SMAD4	2,13E-02	downregulation
EPITHELIAL MESENCHYMAL TRANSITION	2,14E-02	downregulation
KLFS	2,42E-02	downregulation
IL6 JAK STAT3 SIGNALING	2,97E-02	upregulation

- Mitochondria-related pathways
- Pathways known to be dysregulated in CF

Proteomic

PATHWAY	p value	Effect of CFTR modulators
OXIDATIVE PHOSPHORYLATION	2,54E-44	upregulation
FATTY ACID METABOLISM	2,67E-27	upregulation
ADIPOGENESIS	8,42E-18	upregulation
TCA Cycle	3,38E-08	upregulation
PEROXISOME	1,41E-07	upregulation
LPS/IL-1 Mediated Inhibition of RXR Function	3,48E-06	upregulation
Valine Degradation	1,08E-05	upregulation
MITOTIC SPINDLE	6,87E-05	downregulation
Superpathway of Cholesterol Biosynthesis	2,63E-04	upregulation
XENOBIOTIC METABOLISM	2,76E-04	upregulation
Signaling by Rho Family GTPases	8,34E-04	downregulation
Stearate Biosynthesis	8,34E-04	upregulation
BILE ACID METABOLISM	9,68E-04	upregulation
FOX P1	9,88E-04	downregulation
INTEGRIN1 PATHWAY	1,42E-03	downregulation
MEF2A	1,52E-03	downregulation
TEAD1	1,83E-03	downregulation
G2M CHECKPOINT	2,03E-03	downregulation
GRHL2	2,30E-03	downregulation
POU5F1	2,83E-03	downregulation
MTORC1 SIGNALING	4,31E-03	upregulation
EPITHELIAL MESENCHYMAL TRANSITION	5,77E-03	downregulation
REACTIVE OXYGEN SPECIES PATHWAY	5,89E-03	upregulation
Glycolysis	6,73E-03	upregulation
Leukocyte Extravasation Signaling	6,73E-03	downregulation

Figure 29 : Liste des voies différentiellement exprimées en transcriptomique et en protéomique chez les patients homozygotes pour la mutation F508del selon qu'ils soient ou non traités par VX-770/VX-661/VX-445. Rouge : Voies déjà connues comme étant dérégulées dans la mucoviscidose.

Vert : voies liées aux mitochondries.

Dans le cas de l'étude des patients porteurs de mutations de classe I, seule une analyse en protéomique a été réalisée. 8751 protéines ont ainsi été détectées et conservées. Les patients homozygotes pour la mutation G542X ont dans un premier temps été considérés séparément des patients porteurs d'autres mutations de classe I. Cependant le trop faible nombre d'individus n'a pas permis de mettre en évidence des protéines ou des voies de signalisation spécifiques à ces sous-groupes vis-à-vis du traitement par modulateurs. Dans un second temps les 9 patients ont été considérés ensemble. Seulement 7 protéines ont été détectées comme étant différentiellement exprimées entre les cellules traitées au DMSO et celles traitées par la triple combinaison (SUOX, OSBPL5, MIPEP, CYP51A1, VWA8, DAP3, SOAT1), mais 18 voies sont apparues différentiellement exprimées après analyse par fGSEA, les voies les plus significatives comprenant là encore de manière intéressante les voies liées aux mitochondries et au cycle de Krebs. rRoma confirme ces observations, l'algorithme mettant en lumière plusieurs voies liées aux mitochondries. De manière intéressante, l'analyse rRoma montre aussi que ces voies sont plus corrigées par les modulateurs chez les patients porteurs de la mutation G542X que chez les patients porteurs d'autres

mutations de classe I. Le détail des voies obtenues via fGSEA et rRoma est présenté figure 29 et en annexe 10.

3.4 Discussion

Nos résultats révèlent un rôle important des modulateurs de CFTR dans la correction des voies mitochondriales, ce qui est mis en avant à la fois en transcriptomique et en protéomique, et ce au moyen de différents outils d'analyse, ce qui indique une robustesse de ces résultats. De manière intéressante, cette correction est retrouvée à la fois chez les patients porteurs de mutations de classe II et chez ceux porteurs de mutations de classe I, laissant supposer que cet effet de correction serait indépendant de CFTR.

Or, les mitochondries jouent un rôle clé dans les cellules, notamment dans l'apoptose, l'homéostasie du calcium et la production d'espèce réactives de l'oxygène (Duchen 2004, Murphy 2009, Murphy 2009), mais aussi et surtout pour la production d'énergie pour la cellule en produisant l'ATP.

La mitochondrie est un organite semi-autonome doté d'un génome propre (Scarpulla 2008). C'est à l'intérieur de la mitochondrie qu'ont lieu les réactions enzymatiques du cycle de Krebs, de la production du NADH et du FADH₂ qui agissent comme porteurs d'électrons vers les complexes de la chaîne respiratoire (elle aussi présente dans la mitochondrie), à la consommation d'oxygène permettant de récupérer l'énergie provenant des molécules à haute énergie issues de l'alimentation pour produire de l'ATP. Tout défaut dans le flux énergétique altère l'homéostasie mitochondriale et induit des conditions pathologiques. En effet, l'énergie extraite au niveau des mitochondries au cours du cycle de Krebs permet non seulement de fournir à la cellule de l'ATP nouvellement synthétisé, mais elle est aussi une caractéristique cruciale des mitochondries saines (Mitchell 1961), étant la force motrice d'autres processus mitochondriaux, tels que les importations de protéines mitochondriales (Neupert 2007) ou tout autre changement de comportement des mitochondries en réponse à un dysfonctionnement. En outre, la mitochondrie est au cœur du métabolisme. L'oxydation des acides gras peut notamment y avoir lieu.

Une conséquence de la respiration mitochondriale est la génération d'électrons non appariés. L'oxygène moléculaire peut être réduit en H₂O par un seul électron à la fois, mais il peut arriver que des électrons non appariés réduisent O₂ pour produire l'anion superoxyde (O₂⁻) (Mailloux 2011). O₂⁻ et H₂O₂ (aussi produit lors de la respiration mitochondriale) sont deux espèces réactives de

l'oxygène. Ainsi, des dysfonctionnements au niveau mitochondrial peuvent être liés à une augmentation de la production d'espèces réactives de l'oxygène, et donc à une augmentation du stress oxydatif de la cellule. Pourtant, notre étude révèle que des voies à l'origine de la production d'espèces réactives de l'oxygène dans les mitochondries sont encore plus fortement exprimées en réponse au traitement par modulateurs. Deux hypothèses sont alors envisageables : soit les modulateurs participent au dérèglement de l'homéostasie des mitochondries, ce qui semble peu probable au vu de leur effet bénéfique sur les patients atteints par la mucoviscidose, soit ils améliorent l'efficacité de la respiration mitochondriale en jouant sur le fonctionnement des mitochondries ou sur leur quantité, ce qui mènerait mécaniquement à une augmentation de la production d'espèces réactives de l'oxygène.

En plus d'être la principale source des espèces réactives de l'oxygène, les mitochondries contiennent également les défenses antioxydantes de la cellule (Mailloux 2018), qui limitent leurs effets néfastes, protégeant ainsi la cellule. Cela fait de la mitochondrie un acteur central de l'homéostasie oxydoréductrice cellulaire. Des dysfonctionnements subtils la balance oxydo-réductrice de la mitochondrie pourraient donc être à l'origine de dysfonctionnements bien plus importants au niveau cellulaire. Ces défenses antioxydantes sont par ailleurs aussi sur-exprimées dans notre étude en réponse aux traitements par modulateurs, potentiellement du fait de l'augmentation de la production d'espèces réactives de l'oxygène via la suractivation de voies telles que la phosphorylation oxydative.

L'hypothèse selon laquelle les anomalies mitochondriales pourraient être liées d'une manière ou d'une autre aux symptômes retrouvés dans la mucoviscidose a été émise pour la première fois en 1979, lorsque des anomalies de la chaîne respiratoire des mitochondries ont été signalées. Dans cette étude, Shapiro et ses collaborateurs ont constaté que la consommation d'oxygène était plus importante dans les cellules provenant de patients atteints par la mucoviscidose (Favia 2016). De manière cohérente, dans une autre étude, une augmentation d'environ 50 % de la consommation d'oxygène a été décrite dans les cellules épithéliales de patients atteints par la mucoviscidose par rapport à des contrôles sains. Par conséquent, la production mitochondriale de O_2^- et de peroxyde H_2O_2 pourrait également augmenter en raison de l'augmentation de la consommation d'oxygène par les cellules produisant CFTR muté (Turrens 1982).

D'autres anomalies mitochondriales ont été décrites dans les cellules présentant CFTR F508del, telles que la fragmentation du réseau mitochondrial et la réduction de l'absorption mitochondriale de Ca^{2+} , deux événements vraisemblablement liés à une dépolarisation primaire de

la membrane mitochondriale (Valdivieso 2013). Ces résultats semblent indiquer une implication de la mutation de CFTR dans l'altération de la structure et de la fonction mitochondriale.

Par la suite, plusieurs études ont décrit d'autres modifications mitochondriales dans la mucoviscidose (Valdivieso 2013). Cependant, lorsque CFTR a été cloné et identifié comme un canal ions chlorure, les hypothèses d'une possible implication mitochondriale dans la mucoviscidose ont été totalement écartées au point de concentrer toute l'étude sur la protéine mutée. Par conséquent, les travaux ultérieurs ont porté principalement sur CFTR en tant que canal ions chlorure. Seules quelques études ont continué à explorer l'implication des mitochondries dans la mucoviscidose.

Pourtant, en plus d'être un canal ions chlorure, CFTR peut également affecter indirectement l'expression des gènes. Il a déjà été montré que certains gènes dépendant de CFTR sont impliqués dans des voies cellulaires spécifiques liées au métabolisme ou à l'inflammation (Aldini 2018), voies que l'on retrouve par ailleurs majoritairement corrigées par VX-770/VX-661/VX-445 dans nos analyses (on retrouve notamment la voie de signalisation du TNFalpha comme étant sous-exprimée en réponse au traitement) . De plus, deux gènes en particulier, MT-ND4 et CISD1, que l'on retrouve aussi dans nos analyses, et codant pour des protéines mitochondriales, ont déjà été signalés comme étant sous exprimés dans des cellules de patients atteints par la mucoviscidose (Valdivieso 2013).

En particulier, le gène MT-ND4 est essentiel à l'activité correcte de la chaîne respiratoire des mitochondries (Valdivieso 2013). Par conséquent, la régulation négative de la protéine MT-ND4 détectée dans les cellules de patients atteints par la mucoviscidose pourrait être responsable de la faible efficacité de l'oxydation du NADH, et de l'augmentation de la production d'espèces réactives de l'oxygène. Or de manière très intéressante, MT-ND4 est retrouvée comme étant sur-exprimée dans les cellules traitées par VX-770/VX-661/VX-445 dans notre analyse transcriptomique des cellules provenant de patients homozygotes pour la mutation F508del. Cela semble donc indiquer que VX-770/VX-661/VX-445 corrige un dysfonctionnement mitochondrial présent dans les mutations de classe II. Si le gène n'a pas été retrouvé en protéomique, que ce soit chez les patients porteurs de mutations de classe II ou I, la similarité dans les voies retrouvées laisse sous-entendre qu'une correction similaire a aussi lieu pour les mutation de classe I, et donc que le mécanisme de sauvetage des mitochondries serait aussi au moins pour partie indépendant de CFTR, hypothèse qui n'avait jusqu'à présent jamais été formulée. Mais l'effet exact de ces modulateurs sur la mitochondrie reste difficile à établir à partir de notre seule étude : comme évoqué précédemment, malgré une hausse de la quantité de MT-ND4 en réponse au traitement par modulateurs, les voies participant à la production d'espèces réactives de l'oxygène sont elles aussi sur-exprimées.

D'autres études ont par ailleurs montré que la production accrue d'espèces réactives de l'oxygène était cruciale pour la progression de la mucoviscidose (Valdivieso 2013) et, de manière cohérente, des niveaux élevés de produits d'oxydation des lipides et des protéines ont été trouvés dans le liquide de lavage broncho-alvéolaire de patients atteints par la mucoviscidose (Kelly-Aubert 2011). Or, parmi les voies les plus fortement sur-exprimées en présence de VX-770/VX-661/VX-445 dans notre étude, on retrouve la phosphorylation oxydative ou encore le peroxydase, deux voies fortement associées à la production d'espèces réactives de l'oxygène. De plus, ces voies sont retrouvées à la fois chez les patients porteurs de mutations de classe I et II, laissant supposer là aussi que le mécanisme de surproduction d'espèces réactives de l'oxygène est au moins pour partie indépendant de CFTR. VX-770/VX-661/VX-445 semble donc agir à la fois de manière bénéfique pour la cellule en boostant la production d'énergie via une augmentation de l'efficacité de la respiration cellulaire réalisée par les mitochondries, mais aussi potentiellement de manière négative en accélérant la production d'espèces réactives de l'oxygène, et en augmentant donc le stress oxydatif.

3.5 Conclusion

Il est clair que les mitochondries ont un énorme potentiel d'influence sur la santé des patients. Or les résultats de notre étude montrent aussi clairement que la modification du phénotype des cellules des patients porteurs de mutations de classe II via le traitement par VX-770/VX-661/VX-445 est au moins pour partie due à la correction de dysfonctionnements mitochondriaux. En outre, la même correction est observée chez les patients porteurs de mutations de classe I, laissant supposer que les modulateurs de CFTR pourraient être bénéfiques aux patients porteurs de ces mutations, et ce, malgré paradoxalement leur absence d'effet sur CFTR dans le cas de telles mutations. De plus, nos résultats soulignent l'importance des mitochondries dans la mucoviscidose, qui ont été malheureusement très peu étudiées dans le contexte de cette maladie au cours des 20 dernières années. Enfin, ces résultats mènent à penser que l'optimisation du métabolisme des mitochondries dans la mucoviscidose pourrait être au centre d'un traitement thérapeutique efficace. Des études biologiques supplémentaires sont néanmoins nécessaires afin de confirmer les résultats obtenus dans ces études omiques, et sont en cours de réalisation au moment de l'écriture de ce manuscrit. De plus, la question de la production d'espèces réactives de l'oxygène et de leur

potentiel effet délétère pour la cellule en réponse à un traitement par modulateurs doit encore être étudiée plus en détails.

4. rRoma, un algorithme pour l'étude de groupement des gènes lors d'analyses omiques

Un article correspondant à cette section a été soumis à Bioinformatics au moment de la rédaction de ce manuscrit. Un preprint est disponible et donné en annexe.

4.1 Introduction

L'utilisation de techniques génomiques à haut débit telles que la protéomique ou la transcriptomique devient de plus en plus facilitée par l'amélioration des outils d'acquisition de ces données, menant à une baisse drastique des coûts associés à de telles analyses. Ces dernières permettent de mesurer avec précision les profils moléculaires de systèmes biologiques, et ce à plusieurs niveaux (Hawkins 2010). Cependant, et comme l'a montré l'étude précédente, la quantité de données produites lors de telles expériences est très importante, et requiert l'utilisation de logiciels et d'algorithmes dédiés pour les analyser. En particulier, la capacité d'interprétation des résultats en termes d'explication des voies cellulaires impliquées devient une question cruciale. Des analyses dédiées à cette question sont nécessaires pour synthétiser et transformer les données en informations biologiques précieuses (Hawkins 2010).

De nombreuses applications biologiques et cliniques, telles que celles décrites lors des sections précédentes de ce manuscrit, nécessitent la comparaison d'échantillons provenant de conditions différentes. L'objectif d'interprétabilité recherché dans de telles analyses implique souvent de mettre en évidence les processus biologiques qui présenteraient des activités différentes entre les conditions comparées (métabolisme, transcription, traduction, voies de signalisation, etc). Une approche largement utilisée en génomique, et appliquée dans nos études précédentes, consiste à comparer des mesures au niveau d'un seul gène ou d'une seule protéine afin d'identifier des indicateurs caractéristiques d'un état pathologique particulier (des biomarqueurs) ou des gènes moteurs liés de manière causale à l'initiation et à la progression de la maladie étudiée (Barillot et al., 2012).

Une voie de signalisation est généralement définie comme un ensemble de gènes ou de protéines qui interagissent dans l'objectif de transmettre des informations de l'extérieur vers l'intérieur d'une cellule, conduisant à une réponse adaptée de la cellule, c'est-à-dire à l'expression des gènes pertinents compte tenu des informations concernant l'environnement dans lequel se trouve la cellule. Dans de nombreuses maladies, certaines voies sont altérées et les dérégulations

peuvent provenir d'un ou plusieurs gènes de cette voie, qui peuvent différer d'un patient à l'autre. La quantification de l'activité des voies à partir des mesures protéomiques ou transcriptomiques de l'expression de gènes et de protéines est aujourd'hui largement utilisée pour transformer les données au niveau des gènes en données associées à des ensembles de gènes partageant des fonctions similaires, représentant des processus biologiques ou des voies qui peuvent révéler l'hétérogénéité biologique entre les échantillons (Levine et al., 2006 ; Ramos-Rodriguez et al., 2012 ; Borisov et al., 2014).

Ainsi, il est devenu apparent ces dernières années que dans le cancer (et d'autres maladies systémiques), les mêmes voies peuvent être affectées par des défauts dans différents gènes chez des patients différents, et que les profils moléculaires des échantillons de tumeurs sont plus similaires au niveau des voies qu'au niveau des gènes (Wang et al., 2010). L'application d'approches basées sur les voies dans l'analyse des données génomiques peut ainsi aider à capturer des informations biologiques qui sont autrement indétectables en se concentrant uniquement sur les gènes au niveau individuel.

De telles approches pourraient aussi s'avérer intéressantes dans des maladies monogéniques telles que la mucoviscidose. Ici, bien qu'un seul gène soit muté, l'absence de la protéine associée dans les cellules épithéliales est à l'origine d'un grand nombre de phénotypes différents, qui ne semblent pas en lien avec la fonction canal ions chlorure de CFTR. En effet, CFTR fait en réalité partie d'un réseau de protéines complexe (Amaral 2020), dont tous les acteurs n'ont d'ailleurs peut-être pas encore été identifiés. Une hypothèse pour relier CFTR aux nombreux phénotypes de la mucoviscidose serait que le fonctionnement normal de ce réseau serait perturbé par l'absence de cette protéine. Une analyse par voies de signalisation pourrait donc permettre de mieux comprendre les fondements moléculaires de l'ensemble des phénotypes de cette maladie. De plus, dans le cas d'études de l'effet de modulateurs, elle permettrait de mieux comprendre l'ensemble des effets de ces derniers dans les cellules de patients atteints par la mucoviscidose.

Dans cette étude, nous avons développé un algorithme, publié sous forme de package R, rRoma, qui a été conçu pour répondre à la question de la quantification de l'activité d'ensembles de gènes (appelés collections par la suite) caractérisés par leur implication dans un rôle fonctionnel commun pour la cellule. Ces collections peuvent ainsi correspondre à des gènes ayant les mêmes activités fonctionnelles, à des gènes régulés par les mêmes motifs, à des gènes appartenant à la même voie de signalisation, ou encore à des gènes formant un groupe de gènes fréquemment co-exprimés. rRoma est une évolution de ROMA, un programme développé en Java en 2016 (Martignetti 2016)

L'idée derrière rRoma consiste à quantifier l'activité d'une collection en calculant la plus grande quantité de variance unidimensionnelle entre les échantillons expliquée par les gènes de la collection (propriété de la PC1). Cette quantité correspond à l'action d'un facteur caché sur l'expression des gènes de la collection et témoigne de la variabilité de l'activité de ce facteur dans les échantillons étudiés. Ce paramètre permet ensuite de construire le modèle linéaire le plus simple possible pour représenter l'expression des gènes de la collection considérée dans l'échantillon étudié.

Par ailleurs, rRoma met aussi en œuvre plusieurs fonctionnalités supplémentaires. Il permet notamment de déterminer les gènes d'une collection qui contribuent le plus à la définition de la PC1. L'algorithme fournit plusieurs méthodes alternatives pour le calcul de cette dernière, fournissant notamment une version pondérée de l'analyse en composantes principales. Enfin, il estime la signification statistique de la répartition des échantillons le long de la première composante pour une collection de deux manières : il distingue les collections sur-dispersées et les collections déplacées. La sur-dispersion d'une collection de gènes signifie que la quantité de variance expliquée par la PC1 calculée pour un ensemble de données restreint aux gènes cette collection est significativement plus grande que la variance d'un ensemble de gènes choisis aléatoirement et de même taille. Ainsi, la sur-dispersion signifie une plus grande variabilité d'une collection de gènes entre les échantillons étudiés, même si cette propriété n'aurait pas été détectable en considérant les gènes séparément. Le déplacement d'un ensemble de gènes signifie que l'expression médiane de l'ensemble des gènes de la collection est significativement différente de celle de la totalité des gènes étudiés, c'est-à-dire que la collection de gènes étudiés se comporte de manière singulière dans un moins un des échantillons considérés.

La quantification naïve de l'activité d'une collection consiste souvent à calculer la moyenne ou la médiane de l'expression des gènes de la collection dans un échantillon donné ou, à l'opposé, repose sur un seul gène marqueur de l'activité de la collection. rRoma se démarque de ces approches par le fait qu'il est particulièrement adapté pour modéliser les cas où les différents gènes ne contribuent pas individuellement de manière similaire à l'activité de la collection, comme lorsque certains gènes contribuent de manière plus importante que d'autres pour définir l'activité de la collection (par exemple les facteurs de transcription situés en aval des voies de signalisation). Il s'avère par ailleurs particulièrement intéressant dans les cas où certains gènes sont négativement corrélés avec l'activité de la collection : dans un tel cas, la première composante peut capturer cet effet, alors qu'une simple analyse par moyenne ne montrerait rien.

Plusieurs méthodes de quantification des voies ont déjà été proposées pour récapituler l'activité d'une collection en calculant le premier méta-gène dans la décomposition en valeurs singulières de la matrice d'expression restreinte aux gènes de la collection (Tomfohr et al., 2005). Dans l'étude de Bild et al. (Bild 2006), une stratégie similaire a été exploitée afin de définir l'activité de plusieurs voies liées au cancer sur une grande collection de transcriptomes de cancers humains. Dans l'étude de Fan et al. (Fan 2016), les auteurs ont suggéré la notion de "voie sur-dispersée" dans le cadre de l'analyse transcriptomique unicellulaire, de sorte que la mesure de l'activité d'un ensemble de gènes est quantifiée par la signification statistique de la sur-dispersion expliquée par la première composante principale, calculée pour un ensemble de profils transcriptomiques unicellulaires. D'autres méthodes ont été développées pour estimer les scores d'activité des collections dans des échantillons individuels, comme l'extension à un seul échantillon de GSEA (ssGSEA) (Barbie et al., 2009) ou OncoFinder (Borisov et al., 2014).

Le package pour le logiciel R s'appelle rRoma et est open-source. Il est disponible sur github, à cette adresse : www.github.com/sysbio-curie/rROMA.

4.2 Méthodes

4.2.1 UTILISATION DE L'ACP

L'idée principale de rRoma consiste à considérer que pour une collection donnée de gènes, l'expression d'un gène G dans un échantillon S est proportionnelle à l'activité d'un facteur F , de sorte que l'on puisse écrire :

$$Expression(G, S) = \alpha_G^F Activity_S^F$$

Où α_G^F est le coefficient de réponse du gène G au facteur F , et $Activity_S^F$ est l'activité du facteur F dans l'échantillon S . Ces deux valeurs peuvent être facilement déterminées en considérant la première composante de l'ACP des gènes de la collection étudiée dans l'espace constitué par les échantillons. Dans ce cas, le vecteur $Activity^F$ comprenant les activités des différents échantillons étudiés correspond au premier vecteur propre, c'est-à-dire à la première colonne de la matrice de poids. Le vecteur α^F , qui contient les coefficients de réponses des différents gènes de la collection étudiée, correspond quant à lui à la projection des gènes sur la première composante, c'est-à-dire à la première colonne de la matrice de projection, comme le résume la figure ci-après :

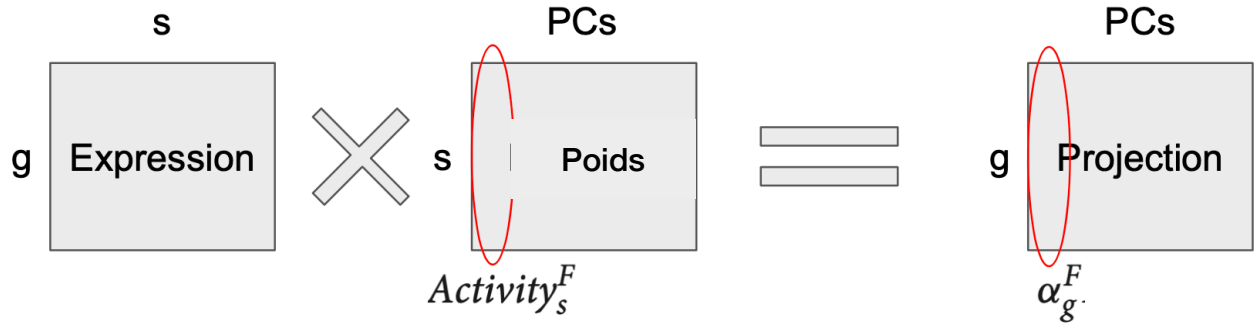


Figure 30 : Représentation des matrices d'intérêt pour la réalisation d'une ACP. Visualisation des données exploitées par rRoma. Le vecteur des activités correspond à la première colonne de la matrice de poids, tandis que le vecteur des coefficients de réponse est la première colonne de la matrice de projection.

Comme pour de nombreuses autres applications de l'ACP, rRoma utilise la décomposition en valeurs singulières afin d'accélérer les calculs. Cependant, pour être valable, l'utilisation de cette méthode nécessite de centrer les données au préalable.

En effet, reprenons les notations définies en introduction. On définit un dataset constitué de n individus, décrits par p variables numériques, que l'on représente par une matrice X de taille $n \times p$. Dans cette matrice, la i ème colonne correspond au vecteur x_i des observations de la variable. Comme expliqué en introduction, l'ACP consiste à déterminer les valeurs propres ainsi que les vecteurs propres de la matrice de variance-covariance de X . Or, cette dernière peut s'écrire $(X - X_{mean})(X - X_{mean})^T / (n - 1)$, où X_{mean} contient les valeurs moyennes des x_i pour chaque colonne.

Dans le cas où les données sont centrées, cette expression se simplifie en $XX^T / (n - 1)$. Dès lors, nous avons une matrice symétrique qui peut donc être diagonalisée. Si l'on note C la matrice de variance covariance de X , il existe L , une matrice diagonale et W telles que $C = WLW^T$.

Revenons à X . D'après le théorème de décomposition en valeurs singulières, il existe une matrice S diagonale, ainsi que deux matrices U et V telles que $X = USV^T$.

On peut donc écrire $C = XX^T / (n - 1) = VSU^TUSV^T / (n - 1) = VS^2V^T / (n - 1)$. Par identification, les vecteurs propres de la matrice de variance covariance de X sont donc donnés par V , tandis que les valeurs propres sont proportionnelles aux valeurs diagonales de S .

Il n'est donc pas nécessaire de calculer la matrice de variance covariance de X , le calcul de la décomposition en valeurs singulières étant équivalent. Ce dernier est par ailleurs bien plus rapide,

et des algorithmes permettent de l'accélérer encore en se concentrant uniquement sur les premières colonnes de V .

rRoma utilise ainsi le package R `irlba` (Baglama 2005), qui permet de se concentrer sur la première composante principale, en ne calculant que la première colonne de V et la première valeur de S .

Cependant, si les données ne sont pas centrées, la simplification $C = XX'/(n - 1)$ n'est plus valable, et il faut désormais prendre en compte le fait que la moyenne soit non nulle.

L'algorithme rRoma débute donc en centrant les données au niveau général. Cependant, elles ne sont pas centrées à nouveau sur chaque collection de gènes prise en considération. Ceci implique de prendre des précautions supplémentaires pour s'assurer de la validité de la décomposition en valeurs singulières, que nous détaillerons dans une section suivante.

4.2.2 PRETRAITEMENT DES DONNEES POUR RROMA

Le format d'entrée pour l'expression des gènes ou des protéines pour rRoma est un fichier texte délimité par des tabulations avec des colonnes correspondant aux échantillons biologiques, et des lignes correspondant aux gènes ou aux protéines. La première ligne est supposée contenir les identifiants des échantillons tandis que la première colonne est supposée contenir les noms non redondants des gènes ou des protéines. Si le tableau de données contient des valeurs manquantes, elles peuvent être imputées en utilisant une approximation de la matrice de données avec valeurs manquantes par une matrice complète de rang inférieur. Pour cela, l'utilisateur doit spécifier le rang de la matrice complète approximative qu'il veut utiliser. Ensuite, les composantes principales sont calculées jusqu'au rang spécifié à l'aide d'un algorithme capable de travailler avec des données manquantes (Gorban 2009). Cette décomposition en ACP est utilisée pour construire la matrice approximative complète de rang inférieur, à partir de laquelle les valeurs manquantes des données initiales sont imputées. Dans la suite de l'algorithme, c'est la matrice imputée complète qui est utilisée.

4.2.3 ORIENTATION DES COMPOSANTES

Dans l'ACP standard, toutes les composantes sont calculées avec un signe d'orientation indéfini : il y a une symétrie miroir, ce qui fait qu'il est difficile de déterminer si une collection donnée de gènes est sur- ou sous- activée. Dans rRoma, plusieurs méthodes existent pour tenter de résoudre ce problème.

Si des connaissances existent sur le rôle d'un gène dans une collection donnée, nous recommandons de les utiliser en associant un signe à l'effet du gène dans la collection : négatif pour un inhibiteur, et positif pour un activateur par exemple. rRoma utilise alors l'information sur ces signes pour choisir l'orientation qui maximise le nombre de gènes associés à un signe positif dont la projection dans la PC1 est positive, et le nombre de gènes associés à un signe négatif dont la projection dans la PC1 est négative.

Bien que moins efficaces, d'autres méthodes d'orientation de la première composante existent dans le cas où il n'y a pas de connaissance à priori sur les gènes dans les collections étudiées. La plus efficace consiste à ne considérer que les gènes associés aux valeurs les plus extrêmes selon la première composante (selon un pourcentage défini par un hyperparamètre modifiable), puis à sommer ces valeurs. Si le résultat est négatif, alors l'orientation de la PC1 est inversée. La logique de cette méthode est d'orienter la première composante selon les gènes les plus importants pour la collection étudiée.

4.2.4 RETRAIT DES ECHANTILLONS « ANORMAUX »

Les mesures peuvent avoir été réalisées de manière incorrecte dans certains échantillons, et les conserver peut conduire à des résultats erronés de rRoma. Par défaut dans l'algorithme, aucun filtrage des échantillons n'est effectué, car la matrice utilisée en entrée est supposée ne contenir que des échantillons corrects, et qui peuvent être pris en considération dans les analyses.

Cependant, la vérification des échantillons peut être activée par un hyperparamètre, et 2 étapes de filtrage sont alors effectuées. Nous nous assurons d'abord qu'un nombre similaire de gènes est détecté dans tous les échantillons, le seuil de différence étant défini manuellement. Ensuite, les échantillons sont projetés dans l'espace des gènes, et nous nous assurons qu'aucun échantillon n'est trop éloigné des autres. Le nombre de PCs utilisées pour effectuer le filtrage à cette étape et le seuil de différence autorisé sont définis par deux hyperparamètres. Il est aussi possible de ne réaliser qu'une seule des deux étapes de filtrage.

4.2.5 SUPPRESSION DES VALEURS ABERRANTES

Le calcul de la PC1 peut se retrouver affecté par la simple présence d'une valeur aberrante dans l'ensemble des données. Cette dernière pourrait en effet orienter artificiellement la PC1 dans sa direction. Afin d'augmenter la robustesse du calcul de la PC1, nous utilisons dans rRoma l'approche

de validation croisée "leave-one-out" (Hastie et al., 2001). Cette méthode fonctionne de la manière suivante :

Dans un premier temps, pour chaque gène du jeu de données considéré, nous calculons le pourcentage de variance expliqué par la seule PC1 lorsque l'on retire le gène du jeu de données. Chaque gène est alors associé à un pourcentage.

Dans un second temps, la distribution de ces pourcentages est centrée et réduite pour obtenir des z-scores.

Dans un troisième temps, tous les gènes dont le z-score associé est au-delà d'une valeur seuil fixée dans l'algorithme (à 3 par défaut, mais modifiable) sont considérés comme aberrants et retirés de l'analyse.

L'idée derrière cette méthode est d'identifier des gènes ayant un impact trop important sur l'ACP à eux seul : si le pourcentage de variance expliqué par la PC1 augmente significativement en l'absence d'un seul gène, cela signifie que ce dernier ne suit pas la logique d'alignement de tous les autres. Il est aberrant et ne devrait pas être pris en considération.

Lorsqu'un gène est considéré comme aberrant pour une collection donnée, il n'est retiré que de celle-ci (mais ce comportement par défaut peut être modifié par un hyperparamètre pour que les gènes soient totalement retirés de toutes les analyses, dès lors qu'ils sont considérés comme aberrants dans au moins une collection). Cependant, dans certains cas, ces gènes peuvent tout de même se révéler intéressants pour l'analyse. Des étapes supplémentaires d'analyse de ces derniers sont donc disponibles dans rRoma.

En particulier, plus le nombre de collections analysées contenant un gène donné est important, plus il est probable que ce dernier soit considéré comme aberrant dans une collection au minimum, et soit donc retiré de l'analyse. Afin d'éviter de tels retraits abusifs, un test de Fischer est réalisé. Celui-ci consiste à comparer la proportion moyenne de collections dans lesquelles les gènes de l'analyse sont considérés comme aberrants à cette même proportion pour un gène en particulier. Si la proportion d'aberrations est proche de la proportion moyenne pour tous les gènes, alors il n'est plus considéré comme aberrant. A l'inverse, si la proportion d'aberrations est significativement plus élevée (seuil déterminé par un hyperparamètre modifiable, à 0.05 par défaut), alors il reste considéré comme aberrant.

A l'inverse, un gène présent uniquement dans un petit nombre de collections peut en réalité être important pour la compréhension de ces dernières. Dans de tels cas, il n'est pas souhaitable de le retirer, même s'il est considéré comme aberrant. Ainsi, si un gène est présent dans moins d'un

nombre défini de collections (fixé par un hyperparamètre et modifiable), alors il n'est pas considéré comme aberrant, et ce, quels que soient les résultats de l'approche « leave-one-out » pour celui-ci.

Enfin, il est possible qu'une proportion importante de gènes soient considérés comme aberrants pour une collection donnée. Or retirer un trop grand nombre de gènes peut totalement dénaturer le comportement détectable pour une collection. Un dernier filtre existe donc pour limiter la proportion maximale de gènes qui peuvent être considérés comme aberrant pour une collection donnée et retirés de l'analyse (la proportion est fixée par un hyperparamètre modifiable). De cette manière, seuls les gènes présentant les variations par approche « leave-one-out » les plus extrêmes sont effectivement considérés comme aberrants.

4.2.6 INTERPRETATION DES RESULTATS

Après avoir réalisé toutes les étapes de pré-processing du jeu de données, impliquant d'imputer les valeurs manquantes, de retirer les échantillons aberrants, de centrer les données au niveau global, puis de retirer de chaque collection les gènes aberrants, le coeur de l'algorithme rRoma peut débuter. Chaque collection est alors considérée séparément.

L'algorithme ne considère que les gènes de la collection en cours d'analyse. Il calcule dans un premier temps la première composante de ces gènes dans l'espace des échantillons. Deux mesures d'intérêt sont alors prises en compte : le pourcentage de variance expliqué par la première composante seule, et la valeur d'expression moyenne des gènes projetés sur cette première composante.

Les mêmes mesures sont ensuite réalisées pour une centaine de collections de gènes générées aléatoirement, qui constitueront la distribution nulle de référence.

Comme expliqué dans la section introductive sur l'ACP appliquée à rRoma, l'utilisation de la décomposition en valeurs singulières n'est justifiée que si les données sont centrées. Si cette condition n'est pas respectée, la projection moyenne sur la première composante ne sera pas nulle. Pour cette raison, la valeur moyenne de la projection des gènes de la collection étudiée sur la PC1 est comparée aux valeurs obtenues pour les collections aléatoires de gènes qui constituent la distribution nulle et sont donc supposées être centrées. Etant donné que ces valeurs peuvent être positives ou négatives, mais que l'on s'intéresse à la déviation par rapport au centre, ce sont les valeurs absolues qui sont prises en considération. Si moins de 5% (valeur définie par un hyperparamètre modifiable) des valeurs obtenues dans la distribution nulle sont inférieures à celle obtenue pour la collection de gènes étudiée, alors cette dernière est dite « déplacée » (ppv Median

Exp <0.05). Cela signifie que l'expression moyenne des gènes dans la collection est différente de l'expression moyenne de tous les gènes, et ce pour au moins un échantillon. Dans un tel cas, des analyses supplémentaires sont nécessaires pour déterminer précisément l'origine de ce déplacement.

Si la collection n'est pas déplacée, alors l'hypothèse des données centrées est considérée comme valable. On s'intéresse alors au pourcentage de variance expliqué par la première composante. Si celui de la collection de gènes étudiée est supérieur à au moins 95% (valeur définie par un hyperparamètre modifiable) des valeurs obtenues pour la distribution nulle, alors la collection en question est considérée comme « surdispersée » (ppv L1 < 0.05). L'hypothèse du facteur unique décrite dans la partie explicative sur l'utilisation de l'ACP dans rRoma est valable pour cette collection de gènes.

ANALYSE D'UNE COLLECTION DE GENES DEPLACEE

Pour les collections déplacées, la mesure de l'activité des échantillons est particulièrement importante. En effet, elle détermine quels échantillons sont responsables du déplacement de la collection.

Si les échantillons étaient déjà séparés en plusieurs groupes avant de l'analyse, il est possible de vérifier que cette séparation en groupes est bien responsable du déplacement, en vérifiant que les scores d'activité sont significativement différents entre les groupes.

A l'inverse, si les groupes ne sont pas connus en amont, il est possible d'en déterminer de nouveaux en réalisant une analyse de groupement hiérarchique sur les collections qui sont déplacées, et donc de déterminer de potentiels nouveaux groupes d'intérêt pour l'analyse en cours.

ANALYSE D'UNE COLLECTION DE GENES SUR-DISPERSEE

Les analyses évoquées précédemment pour le cas des collections déplacées le sont aussi pour les collections sur-dispersées. Mais dans le cas de ces dernières, l'analyse du poids des gènes devient aussi intéressante.

Les gènes associés aux poids les plus forts sont moteurs dans l'activité de la collection. Ils permettent ainsi de résumer l'information des collections à quelques gènes, ce qui peut s'avérer particulièrement utile, notamment pour des interprétations à l'aide d'approches de biologie des systèmes.

L'analyse du signe du poids des gènes est elle aussi particulièrement intéressante. Ainsi, des collections contenant à la fois des activateurs et des inhibiteurs pourront être mises en évidence par rRoma en étant sur-dispersées, et les gènes associés mis en valeur. De telles collections ne seraient

pas détectées comme sur-dispersées par des méthodes d'analyses basées uniquement sur l'expression moyenne des gènes dans les échantillons.

4.2.7 OPTIMISATION DU CALCUL DES DISTRIBUTIONS NULLES

En pratique, il est souvent nécessaire de tester de nombreuses collections de gènes différentes, disponibles dans des bases de données comme KEGG ou Hallmark. L'estimation de la distribution nulle pour chaque ensemble de gènes testé peut conduire à des calculs très coûteux en termes de temps. Dans rRoma, nous ne calculons pas les scores de significativité de la sur-dispersion et du déplacement pour toutes les collections, mais nous les approximations sur une grille prédéfinie de valeurs dépendant de la taille des collections étudiées. En effet, ces deux valeurs étant dépendantes de la taille de l'échantillon considéré, il n'est pas possible d'utiliser la même distribution nulle pour toutes les collections de gènes. Afin d'estimer rapidement l'importance des scores de sur-dispersion et de déplacement, nous construisons donc les distributions nulles pour une liste représentative de tailles de collections. Les tailles de collections représentatives sont choisies pour être uniformément distribuées dans l'échelle logarithmique entre la taille minimale et la taille maximale des collections étudiées. Pour une collection donnée, on choisit ensuite la distribution nulle qui est la plus proche en terme de taille dans l'échelle logarithmique.

4.3 Application pratique

Nous avons utilisé rRoma pour étudier l'activité de voies de signalisation dans les cellules épithéliales des voies respiratoires de patients atteints par la mucoviscidose (N=6) et de donneurs sains (N=6). Nous avons plus précisément comparé les transcriptomes issus de cultures primaires de cellules épithéliales des voies respiratoires de patients porteurs de la mutation la plus fréquente causant la mucoviscidose, F508del, avec ceux de témoins sains, en utilisant des données RNA-seq accessibles en libre accès dans la base de données GEO du NCBI, sous l'identifiant d'accès GSE 176121 (Rehman et al, 2021).

Pour cette analyse, nous avons sélectionné une base de données de collections de gènes, Hallmark, issue de la Molecular Signature Database (MsigDB)(Liberzon et al, 2015). Il s'agit d'une base de données constituée de cinquante collections de gènes, c'est à dire d'un nombre restreint de collections. Chaque collection correspond à un état ou à un processus biologique bien défini. Cette base de données présente par ailleurs l'avantage d'être publique et en libre accès.

Cependant, pour réaliser une analyse plus complète des processus biologiques potentiellement impliqués dans une maladie considérée, rRoma peut aussi être appliqué en utilisant plusieurs bases de données de voies de signalisations combinées ensemble. Lors de l'application de rRoma pour une étude omique, nous recommandons donc de prendre en considération le plus de voies de signalisation possibles, tout en restant cohérent vis-à-vis de l'étude menée, et sans que le temps nécessaire à l'application de rRoma ne devienne prohibitif.

rRoma est ensuite exécuté en spécifiant simplement la base de données de voies à utiliser et la matrice d'expression à analyser, comme indiqué dans la section précédente.

Les voies activées sont déterminées à partir de la matrice appelée ModuleMatrix calculée par rRoma. Comme expliqué dans la section précédente, Les voies dont la p-value associée à l'expression médiane est inférieure à un certain seuil sont considérées comme déplacées, tandis que celles dont la p-value L1 (p-value pour la sur-dispersion) est inférieure à ce même seuil sont considérées sur-dispersées. L'activité des voies dans les différents échantillons peut être représentée graphiquement avec la fonction Plot.Genesets.Samples (figures 31 et 32), et les gènes contribuant le plus à chaque voie sont déterminés en visualisant les poids des gènes, grâce à la fonction GetTopContrib (figure 33). Des diagrammes en boîte à moustaches représentant les scores d'activité dans des groupes d'échantillons prédéfinis peuvent également être tracés, et permettent une analyse différentielle. Dans notre exemple, toutes les voies mises en évidence se comportent de manière significativement différente chez les patients atteints de mucoviscidose et chez les donneurs sains (figure 34).

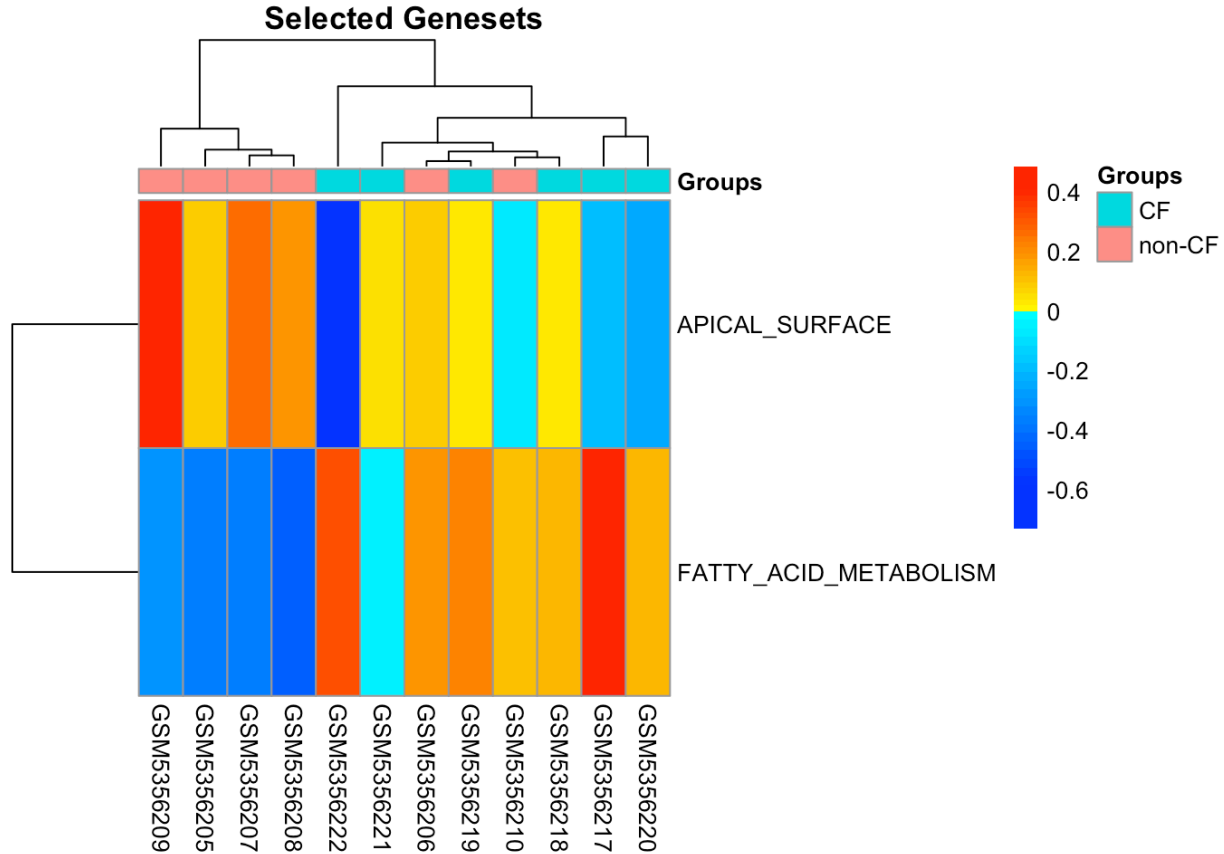


Figure 31 : Activité des échantillons pour les deux voies déplacées. Les activités les plus élevées sont en rouges et les plus faibles en bleu. Un dendrogramme est tracé pour regrouper les patients selon leur niveau d'expression pour les collections considérées.

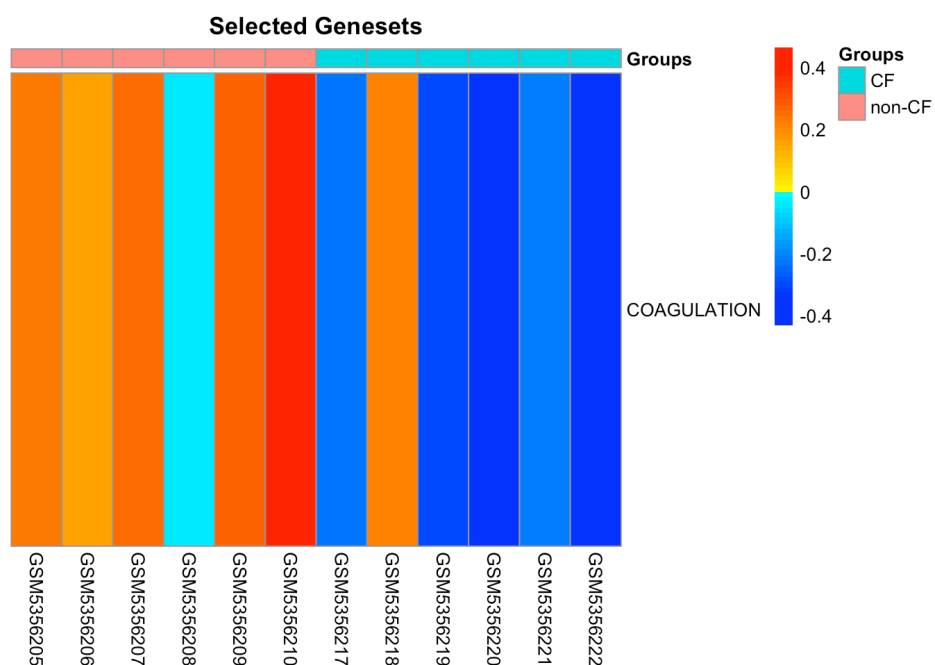


Figure 32 : Activité des échantillons pour la voie surdispersée. Les activités les plus élevées sont en rouges et les plus faibles en bleu

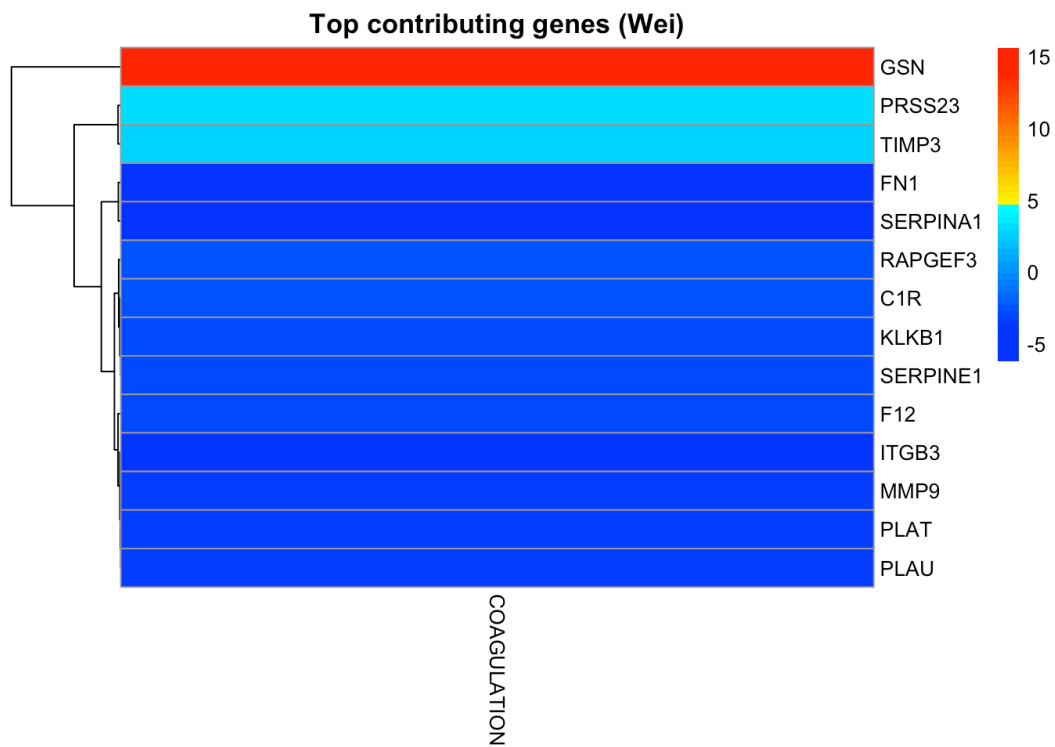


Figure 33 : Gènes aux poids les plus importants pour la voie surdispersée. Les gènes sont triés par niveau d'importance, le plus important se trouvant en haut.

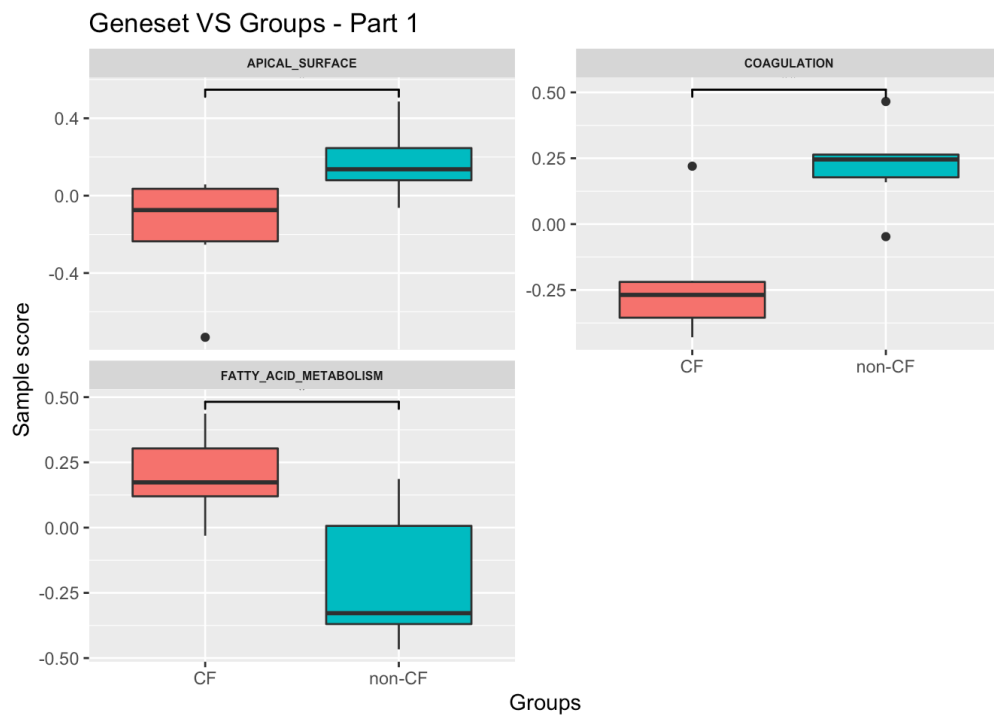


Figure 34 : Comparaison des scores d'activité entre les deux groupes étudiés pour les 3 collections surdispersées ou déplacées.

Dans notre exemple, sur les 50 voies Hallmark testées, 3 étaient significativement actives, et étaient associées au métabolisme des acides gras, à la surface apicale et la coagulation. La voie associée au métabolisme des acides gras (HALLMARK_Fatty_acid_metabolism) présente des scores d'activité significativement différents entre les patients atteints de la mucoviscidose et les donneurs sains. Cette voie a été largement étudiée dans le contexte de la mucoviscidose, et sa dérégulation est un phénotype bien connu de la maladie (Strandvik et al 2010). Cela illustre donc bien la capacité de rRoma à extraire les dérégulations biologiquement pertinentes à partir des données transcriptomiques.

La voie liée à la surface apicale est plus difficile à interpréter. Elle pourrait en réalité mettre en évidence des différences survenant au cours de la différenciation cellulaire, et elle pourrait donc être liée à la culture cellulaire plutôt qu'à la maladie. En effet, les cellules considérées dans cette études ont cultivées sur filtres, de manière similaire à ce qui a été décrit dans la section précédente. Les cellules ont ainsi été cultivées à l'interface air-liquide durant plusieurs semaines afin d'être différenciées. De légères modifications des conditions de culture durant cette période de différenciation auraient pu altérer le profil apical des cellules, et donc modifier l'expression du groupement de gènes associé.

Enfin, la voie liée à la coagulation, la seule voie sur-dispersée, semble être liée à un gène spécifique avec un poids associé très élevé : la gelsoline (GSN). GSN est en effet de loin le gène qui contribue le plus au score d'activité de cette voie liée à la coagulation. La gelsoline a déjà été mise en évidence comme jouant un rôle dans l'activation de CFTR (Cantiello 2001, Vasconcellos 1994), ce qui indique que la détection de cette voie est compatible avec les connaissances disponibles concernant la mucoviscidose.

Ainsi, dans l'ensemble, dans cette étude, rRoma a mis en évidence un mécanisme pertinent dans le contexte de mucoviscidose, un biais potentiel dû à la culture cellulaire, et un gène intéressant qui pourrait être étudié de manière plus approfondie.

Comme expliqué lors de la section précédente, de nombreux hyperparamètres peuvent être spécifiés et changés pour modifier la vitesse, la précision ou le comportement de rRoma vis-à-vis des valeurs aberrantes. Le temps de calcul nécessaire pour exécuter l'algorithme dépend généralement de ces paramètres (en particulier, de l'activation ou non de la parallélisation) ainsi que du nombre de voies étudiées et de leurs tailles respectives. En ce qui concerne l'exemple discuté ici, l'algorithme s'est exécuté en environ 3 minutes et 15 secondes sur un MacBook Pro équipé d'un processeur Intel Core i7 6 cœurs à 2,6 GHz. L'analyse d'une seule voie de 60 gènes a pris environ 5

secondes. La parallélisation n'a pas été utilisée, mais elle aurait augmenté la vitesse de l'analyse si elle l'avait été.

4.4 Conclusion

En résumé, cette étude montre que rRoma est capable d'identifier des voies génétiques associées aux maladies à partir de données transcriptomiques ou protéomiques, permettant à la fois une interprétation plus claire des résultats d'un point de vue biologique, et une interprétation des changements cellulaires d'une manière plus globale et plus fonctionnelle.

Une vignette détaillée permettant de reproduire toutes les analyses présentées ici est également disponible sur le GitHub contenant le code source de l'algorithme.

Cet algorithme a prouvé son utilité dans le cadre de la mucoviscidose, comme a pu le montrer l'étude décrite dans la section précédente, au cours de laquelle rRoma a été utilisé et a permis de mettre en évidence de nouvelles collections de gènes d'intérêt.

Discussion générale

Au cours de cette thèse, j'ai eu l'opportunité de m'intéresser à de nombreux aspects de la mucoviscidose, en utilisant des méthodes et des compétences variées, faisant appel à des méthodes informatiques, ainsi qu'à des expériences de laboratoire en biologie. Les questions traitées étaient nombreuses, toutes en lien avec les modulateurs de CFTR, dans l'espoir de pouvoir mieux comprendre ces derniers, que ce soit dans leur fonctionnement ou dans la manière d'évaluer leurs effets en clinique. Une meilleure compréhension du mécanisme d'action des modulateurs aurait une très grande utilité dans la lutte contre la mucoviscidose : si les modulateurs s'avéraient bénéfiques pour les patients via des cibles contre lesquelles ils n'ont pas été spécifiquement développés, cela signifierait que de nouveaux médicaments spécifiques de ces cibles pourraient avoir des effets encore plus importants.

L'évaluation de ces bénéfices relève d'un double enjeu. Il existe plusieurs milliers de mutations différentes causant la mucoviscidose, et les effets des modulateurs varient en fonction de ces mutations. Ces dernières ont été regroupées en grandes classes, et les mutations de classe I et VII en particulier apparaissent comme particulièrement importantes du fait de l'absence d'effet des modulateurs sur CFTR pour ces mutations. Cependant, un effet bénéfique des modulateurs chez les patients porteurs de telles mutations n'est pas à exclure. Ce constat a d'ailleurs fait l'objet de la troisième étude décrite dans ce manuscrit.

Le second enjeu tient à la capacité d'évaluer un effet clinique bénéfique des modulateurs chez les patients. Comme évoqué en introduction de ce manuscrit, cette tâche est de plus en plus difficile à mesure que les conditions de vie des patients s'améliorent avec l'apparition de nouveaux traitements. En particulier, il est clair que tous les patients ne répondent pas de la même manière aux traitements, pourtant cette hétérogénéité n'a que très peu été étudiée.

Dans notre première étude, nous avons ainsi pu montrer que dans le cas spécifique de la combinaison VX-770/VX-809 chez des enfants de 6-18 ans, les variations du VEMS, de l'IMC et du nombre de jours d'ATB sont trois paramètres importants pour la caractérisation de l'hétérogénéité de la réponse des enfants. En particulier, l'IMC revêt une importance toute particulière dans l'hétérogénéité de la réponse chez les enfants de 6 à 12 ans dont la fonction respiratoire est normale. Ainsi, nous montrons que la mesure seule du VEMS n'est pas suffisante à la caractérisation d'une réponse aux traitements par modulateurs. Il apparaît donc clairement, à la lueur de ces résultats, que la réponse clinique dans la mucoviscidose ne peut plus se restreindre à

l'évaluation d'un seul paramètre clinique. A l'inverse, plusieurs combinaisons de paramètres peuvent être prises en compte afin de mettre en évidence plusieurs profils de réponse, et ces combinaisons varient en fonction de l'âge. De plus, il est probable que ces combinaisons seraient aussi différentes dans le cas d'un traitement utilisant d'autres combinaisons de modulateurs. Nous n'avons malheureusement pu mener notre étude que chez des patients traités par la combinaison VX-770/VX-809, mais la même méthodologie serait applicable à d'autres combinaisons.

Pour mieux comprendre le mécanisme d'action des modulateurs de CFTR, nous nous sommes concentrés sur des approches nouvelles, faisant sens dans le contexte de la mucoviscidose. Nous avons ainsi étudié la composition de la sueur de patients atteints par la mucoviscidose, ce biofluide étant utilisé lors du diagnostic de la maladie. Il est malheureusement très vite apparu que l'objectif premier de cette étude ne pourrait pas être atteint : les analyses différentielles menées sur des patients traités ou non par modulateurs n'ont pas permis de mettre en évidence de nouvelles protéines ou de nouvelles voies d'intérêt dans le contexte de la mucoviscidose. Cependant les limitations étaient potentiellement d'origine technique, et il serait intéressant d'étudier à nouveau la protéomique de ce biofluide pour d'autres patients traités par modulateurs.

Mais l'analyse de la sueur s'est avérée très intéressante sur un autre aspect : elle nous a permis d'améliorer notre compréhension de la physiologie de la maladie, en mettant en avant des pistes sur l'origine de la KPA chez les patients atteints par la mucoviscidose.

Face au relatif échec de l'analyse de la sueur pour évaluer l'effet des modulateurs, nous avons décidé de nous reconcentrer sur l'analyse de matériel génétique et protéomique issu de cellules épithéliales, mimant le comportement des cellules d'un poumon humain.

Le nombre de patients étant limité dans l'étude que nous souhaitons mener, nous avons besoin d'outils nous assurant d'obtenir des résultats malgré une quantité limitée de données. L'approche que nous avons privilégiée est celle de l'analyse par voies. Cependant peu de logiciels existent pour réaliser une telle analyse, et ils ne permettent pas de prendre en compte des groupes comprenant à la fois des activateurs et des inhibiteurs. Nous avons donc repris le développement de ROMA, un programme dédié à l'analyse de voies, adapté au cas d'inhibiteurs et d'activateurs combinés, et avons ainsi créé un package R, rRoma.

A l'aide de ce programme, couplé à d'autres logiciels d'analyse de voies, notamment GSEA, nous nous sommes alors intéressés plus particulièrement à l'effet de la trithérapie VX-770/VX-661/VX-445, qui n'a encore que très peu été étudiée en omique. Nous avons l'espoir de mettre en évidence des mécanismes de correction indépendants de CFTR, qui pourraient donc être utiles aux mutations de classe I et VII, et pour lesquels de nouvelles thérapies plus ciblées pourraient être

développées. Pour cela, nous avons analysé l'effet de la trithérapie à la fois sur des mutations de classe II et sur des mutations de classe I. Nous avons ainsi pu mettre en évidence un effet bénéfique de la trithérapie sur la correction des dysfonctionnements mitochondriaux. De manière très intéressante, cette correction a été retrouvée à la fois chez les patients porteurs de mutations de classe II et chez ceux porteurs de mutations de classe I. Ces résultats mènent aussi à penser que l'optimisation du métabolisme des mitochondries dans la mucoviscidose pourrait être au centre d'un traitement thérapeutique efficace. Ces dernières pourraient d'ailleurs constituer un potentiel biomarqueur d'intérêt pour évaluer la réponse clinique aux modulateurs de CFTR.

Il est désormais nécessaire de confirmer biologiquement les résultats qui ont été mis en évidence au cours de ces analyses informatiques. En particulier, au moment de la rédaction de ce manuscrit, des analyses en Seahorse ainsi que des immunoblots sont en cours de réalisation dans l'espoir de mettre en évidence une augmentation de l'activité mitochondriale en réponse à un traitement par modulateurs, voire une augmentation de la quantité de mitochondries présentes dans les cellules, ou un changement de leur localisation au sein de ces dernières.

Une telle validation, dans un premier temps chez des patients porteurs de mutations de classe II, puis chez des patients porteurs de mutations de classe I, confirmerait l'importance des mitochondries dans la lutte contre les symptômes causés par la mucoviscidose, et ouvrirait la voie à la recherche de nouveaux traitements contre cette maladie.

Conclusion générale

Cette thèse a permis d'améliorer notre compréhension du fonctionnement et des effets des modulateurs de CFTR, à la fois au niveau clinique et au niveau cellulaire.

Au niveau clinique, nous avons pu montrer que dans le cas de la combinaison VX-770/VX-809 chez des enfants de 6-18 ans, les variations du VEMS, de l'IMC et du nombre de jours d'ATB sont trois paramètres importants pour la caractérisation de l'hétérogénéité de la réponse des enfants. En particulier, l'IMC revêt une importance toute particulière dans l'hétérogénéité de la réponse chez les enfants de 6 à 12 ans dont la fonction respiratoire est normale. Ainsi, nous montrons que la mesure seule du VEMS n'est pas suffisante à la caractérisation d'une réponse aux traitements par modulateurs.

Il semble cependant difficile de trouver un biomarqueur caractéristique d'une telle réponse dans la sueur des patients, comme l'a montré la deuxième étude. Néanmoins, l'analyse omique de ce biofluide apparaît comme particulièrement intéressante pour mieux comprendre la physiopathologie de la maladie. En particulier, notre étude a permis de mettre en avant des pistes sur l'origine de la KPA chez les patients atteints par la mucoviscidose.

Les évolutions et effet liés aux modulateurs de CFTR semblent donc être à chercher au niveau des cellules épithéliales de patients, comme l'a montré la troisième étude. Notre étude a ainsi clairement montré que la modification du phénotype des cellules des patients porteurs de mutation de classe II via le traitement par VX-770/VX-661/VX-445 est au moins pour partie due à la correction de dysfonctionnements mitochondriaux. En outre, la même correction a été observée chez les patients porteurs de mutations de classe I, ce qui laisse supposer que les modulateurs de CFTR pourraient être bénéfiques aux patients porteurs de ces mutations, et ce, malgré paradoxalement leur absence d'effet sur CFTR dans le cas de telles mutations. Ces résultats mènent aussi à penser que l'optimisation du métabolisme des mitochondries dans la mucoviscidose pourrait être au centre d'un traitement thérapeutique efficace. Ces dernières pourraient par ailleurs consister un potentiel biomarqueur d'intérêt pour évaluer la réponse clinique aux modulateurs de CFTR.

Ces résultats ont pu être obtenus grâce à l'utilisation d'outils dédiés à l'analyse de voies. En particulier, l'algorithme rRoma que nous avons développé a permis de confirmer nos découvertes sur le rôle des mitochondries. rRoma est capable d'identifier des voies génétiques associées aux maladies à partir de données transcriptomiques ou protéomiques, permettant à la fois une interprétation plus claire des résultats d'un point de vue biologique, et une interprétation des changements cellulaires d'une manière plus globale et plus fonctionnelle.

Au cours de cette thèse, nous avons donc mis en lumière de nouveaux profils cliniques de réponse aux traitements par modulateurs, amélioré notre compréhension des symptômes de la peau chez les patients atteints par la maladie, révélé l'effet de la trithérapie sur les mitochondries dans les cellules épithéliales respiratoires des patients, et développé un nouvel algorithme dédié à l'analyse de voies pour les études omiques.

Références

Abu-El-Haija M, Ramachandran S, Meyerholz DK, Abu-El-Haija M, Griffin M, Giriappa RL, et al. Pancreatic Damage in Fetal and Newborn Cystic Fibrosis Pigs Involves the Activation of Inflammatory and Remodeling Pathways. *The American Journal of Pathology*. 2012 Aug;181(2):499–507.

Accurso FJ. Treatment of cystic fibrosis in infants. *The Lancet Respiratory Medicine*. 2018 Jul;6(7):483–4.

Accurso FJ, Rowe SM, Clancy JP, Boyle MP, Dunitz JM, Durie PR, et al. Effect of VX-770 in Persons with Cystic Fibrosis and the G551D- *CFTR* Mutation. *N Engl J Med*. 2010 Nov 18;363(21):1991–2003.

Adewole OO, Erhabor GE, Adewole TO, Ojo AO, Oshokoya H, Wolfe LM, et al. Proteomic profiling of eccrine sweat reveals its potential as a diagnostic biofluid for active tuberculosis. *Prot Clin Appl*. 2016 May;10(5):547–53.

Aebersold R, Agar JN, Amster IJ, Baker MS, Bertozzi CR, Boja ES, et al. How many human proteoforms are there? *Nat Chem Biol*. 2018 Mar;14(3):206–14.

Akabas MH. Cystic Fibrosis Transmembrane Conductance Regulator. *Journal of Biological Chemistry*. 2000 Feb;275(6):3729–32.

Al-Nakkash L, Hwang TC. Activation of wild-type and $\Delta F508$ -CFTR by phosphodiesterase inhibitors through cAMP-dependent and -independent mechanisms. *Pflgers Archiv European Journal of Physiology*. 1999 Feb 23;437(4):553–61.

Aldini G, Altomare A, Baron G, Vistoli G, Carini M, Borsani L, et al. N-Acetylcysteine as an antioxidant and disulphide breaking agent: the reasons why. *Free Radical Research*. 2018 Jul 3;52(7):751–62.

Alexopoulos A, Chouliaras G, Kakourou T, Dakoutrou M, Nasi L, Petrocheilou A, et al. Aquagenic wrinkling of the palms after brief immersion to water test as a screening tool for cystic fibrosis diagnosis. *J Eur Acad Dermatol Venereol*. 2021 Aug;35(8):1717–24.

Alonso MJ, Heine-Suñer D, Calvo M, Rosell J, Giménez J, Ramos MD, et al. Spectrum of Mutations in the CFTR Gene in Cystic Fibrosis Patients of Spanish Ancestry. *Ann Human Genet*. 2007 Mar;71(2):194–201.

Amaral MD, Hutt DM, Tomati V, Botelho HM, Pedemonte N. CFTR processing, trafficking and interactions. *Journal of Cystic Fibrosis*. 2020 Mar;19:S33–6.

Anderson MP, Berger HA, Rich DP, Gregory RJ, Smith AE, Welsh MJ. Nucleoside triphosphates are required to open the CFTR chloride channel. *Cell*. 1991 Nov;67(4):775–84.

Andrinopoulou ER, Clancy JP, Szczesniak RD. Multivariate joint modeling to identify markers of growth and lung function decline that predict cystic fibrosis pulmonary exacerbation onset. *BMC Pulm Med*. 2020 Dec;20(1):142.

Arkin LM, Flory JH, Shin DB, Gelfand JM, Treat JR, Allen J, et al. High Prevalence of Aquagenic Wrinkling of the Palms in Patients with Cystic Fibrosis and Association with Measurable Increases in Transepidermal Water Loss: Aquagenic Wrinkling and Cystic Fibrosis. *Pediatric Dermatology*. 2012 Sep;29(5):560–6.

Astaiza-Gómez J. Lagrange Multiplier Tests in Applied Research. *SSRN Journal* [Internet]. 2020 [cited 2022 Jun 28]; Available from: <https://www.ssrn.com/abstract=3669884>

Aurora P, Bush A, Gustafsson P, Oliver C, Wallis C, Price J, et al. Multiple-Breath Washout as a Marker of Lung Disease in Preschool Children with Cystic Fibrosis. *Am J Respir Crit Care Med*. 2005 Feb;171(3):249–56.

Averna M, Stifanese R, De Tullio R, Minicucci L, Cresta F, Palena S, et al. Evidence for alteration of calpain/calpastatin system in PBMC of cystic fibrosis patients. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*. 2011 Dec;1812(12):1649–57.

Baechle D, Flad T, Cansier A, Steffen H, Schitteck B, Tolson J, et al. Cathepsin D Is Present in Human Eccrine Sweat and Involved in the Postsecretory Processing of the Antimicrobial Peptide DCD-1L. *Journal of Biological Chemistry*. 2006 Mar;281(9):5406–15.

Baglama James and Lothar Reichel. Augmented implicitly restarted Lanczos bidiagonalization methods. *SIAM Journal on Scientific Computing*. 2005;19–42.

Baker E, Harris WT, Rowe SM, Rutland SB, Oates GR. Tobacco smoke exposure limits the therapeutic benefit of tezacaftor/ivacaftor in pediatric patients with cystic fibrosis. *Journal of Cystic Fibrosis*. 2021 Jul;20(4):612–7.

Baker LB. Physiology of sweat gland function: The roles of sweating and sweat composition in human health. *Temperature*. 2019 Jul 3;6(3):211–59.

Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature*. 2009 Nov;462(7269):108–12.

Bardin E, Pastor A, Semeraro M, Golec A, Hayes K, Chevalier B, et al. Modulators of CFTR. Updates on clinical development and future directions. *European Journal of Medicinal Chemistry*. 2021 Mar;213:113195.

Barillot E, Calzone L, Hupe P, Vert JP, Zinovyev A. Computational Systems Biology of Cancer [Internet]. 0 ed. CRC Press; 2012 [cited 2022 Oct 29]. Available from: <https://www.taylorfrancis.com/books/9781439831458>

Barry PJ, Jones AM, Webb AK, Horsley AR. Sweat chloride is not a useful marker of clinical response to Ivacaftor. *Thorax*. 2014 Jun;69(6):586–7.

Bartoszewski R, Rab A, Twitty G, Stevenson L, Fortenberry J, Piotrowski A, et al. The Mechanism of Cystic Fibrosis Transmembrane Conductance Regulator Transcriptional Repression during the Unfolded Protein Response. *Journal of Biological Chemistry*. 2008 May;283(18):12154–65.

Bell SC, Barry PJ, De Boeck K, Drevinek P, Elborn JS, Plant BJ, et al. CFTR activity is enhanced by the novel corrector GLPG2222, given with and without ivacaftor in two randomized trials. *Journal of Cystic Fibrosis*. 2019 Sep;18(5):700–7.

Berk DR, Ciliberto HM, Sweet SC, Ferkol TW, Bayliss SJ. Aquagenic Wrinkling of the Palms in Cystic Fibrosis: Comparison With Controls and Genotype-Phenotype Correlations. *Arch Dermatol* [Internet]. 2009 Nov 1 [cited 2022 Oct 29];145(11). Available from: <http://archderm.jamanetwork.com/article.aspx?doi=10.1001/archdermatol.2009.260>

Bhalla M, Turcios N, Aponte V, Jenkins M, Leitman BS, McCauley DI, et al. Cystic fibrosis: scoring system with thin-section CT. *Radiology*. 1991 Jun;179(3):783–8.

Bild AH, Yao G, Chang JT, Wang Q, Potti A, Chasse D, et al. Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature*. 2006 Jan;439(7074):353–7.

Bodewes FAJA, Verkade HJ, Wilschanski M. Gastroenterological endpoints in drug trials for cystic fibrosis: CF GI Clinical Endpoints. *Pediatr Pulmonol*. 2016 Oct;51(S44):S18–22.

Bompadre SG, Sohma Y, Li M, Hwang TC. G551D and G1349D, Two CF-associated Mutations in the Signature Sequences of CFTR, Exhibit Distinct Gating Defects. *Journal of General Physiology*. 2007 Apr 1;129(4):285–98.

Bonnart C, Deraison C, Lacroix M, Uchida Y, Besson C, Robin A, et al. Elastase 2 is expressed in human and mouse epidermis and impairs skin barrier function in Netherton syndrome through filaggrin and lipid misprocessing. *J Clin Invest*. 2010 Mar 1;120(3):871–82.

Bonné S, Gilbert B, Hatzfeld M, Chen X, Green KJ, van Roy F. Defining desmosomal plakophilin-3 interactions. *Journal of Cell Biology*. 2003 Apr 28;161(2):403–16.

Borisov NM, Terekhanova NV, Aliper AM, Venkova LS, Smirnov PY, Roumiantsev S, et al. Signaling pathways activation profiles make better markers of cancer than expression of individual genes. *Oncotarget*. 2014 Oct 30;5(20):10198–205.

- Borisov NM, Terekhanova NV, Aliper AM, Venkova LS, Smirnov PY, Roumiantsev S, et al. Signaling pathways activation profiles make better markers of cancer than expression of individual genes. *Oncotarget*. 2014 Oct 30;5(20):10198–205.
- Borowitz D, Lubarsky B, Wilschanski M, Munck A, Gelfond D, Bodewes F, et al. Nutritional Status Improved in Cystic Fibrosis Patients with the G551D Mutation After Treatment with Ivacaftor. *Dig Dis Sci*. 2016 Jan;61(1):198–207.
- Bortoluzzi CF, Volpi S, D’Orazio C, Tiddens HAWM, Loeve M, Tridello G, et al. Bronchiectases at early chest computed tomography in children with cystic fibrosis are associated with increased risk of subsequent pulmonary exacerbations and chronic pseudomonas infection. *Journal of Cystic Fibrosis*. 2014 Sep;13(5):564–71.
- Botelho HM, Uliyakina I, Awatade NT, Proença MC, Tischer C, Sirianant L, et al. Protein Traffic Disorders: an Effective High-Throughput Fluorescence Microscopy Pipeline for Drug Discovery. *Sci Rep*. 2015 Aug;5(1):9038.
- Boucher RC. Airway Surface Dehydration in Cystic Fibrosis: Pathogenesis and Therapy. *Annu Rev Med*. 2007 Feb 1;58(1):157–70.
- Boyle MP, Bell SC, Konstan MW, McColley SA, Rowe SM, Rietschel E, et al. A CFTR corrector (lumacaftor) and a CFTR potentiator (ivacaftor) for treatment of patients with cystic fibrosis who have a phe508del CFTR mutation: a phase 2 randomised controlled trial. *The Lancet Respiratory Medicine*. 2014 Jul;2(7):527–38.
- Bozoky Z, Krzeminski M, Muhandiram R, Birtley JR, Al-Zahrani A, Thomas PJ, et al. Regulatory R region of the CFTR chloride channel is a dynamic integrator of phospho-dependent intra- and intermolecular interactions. *Proc Natl Acad Sci USA* [Internet]. 2013 Nov 19 [cited 2022 Jun 17];110(47). Available from: <https://pnas.org/doi/full/10.1073/pnas.1315104110>
- Braccia C, Tomati V, Caci E, Pedemonte N, Armirotti A. SWATH label-free proteomics for cystic fibrosis research. *Journal of Cystic Fibrosis*. 2019 Jul;18(4):501–6.
- Britto MT, Kotagal UR, Hornung RW, Atherton HD, Tsevat J, Wilmott RW. Impact of Recent Pulmonary Exacerbations on Quality of Life in Patients With Cystic Fibrosis. *Chest*. 2002 Jan;121(1):64–72.
- Bui S, Masson A, Enaud R, Roditis L, Dournes G, Galode F, et al. Long-Term Outcomes in Real Life of Lumacaftor–Ivacaftor Treatment in Adolescents With Cystic Fibrosis. *Front Pediatr*. 2021 Nov 15;9:744705.
- Burat B, Reynaerts A, Baiwir D, Fléron M, Eppe G, Leal T, et al. Characterization of the Human Eccrine Sweat Proteome—A Focus on the Biological Variability of Individual Sweat Protein Profiles. *IJMS*. 2021 Oct 8;22(19):10871.

- Burgel PR, Munck A, Durieu I, Chiron R, Mely L, Prevotat A, et al. Real-Life Safety and Effectiveness of Lumacaftor–Ivacaftor in Patients with Cystic Fibrosis. *Am J Respir Crit Care Med*. 2020 Jan 15;201(2):188–97.
- Burns K, Clatworthy J, Martin L, Martinon F, Plumpton C, Maschera B, et al. Tollip, a new component of the IL-1RI pathway, links IRAK to the IL-1 receptor. *Nat Cell Biol*. 2000 Jun;2(6):346–51.
- Carlile GW, Robert R, Zhang D, Teske KA, Luo Y, Hanrahan JW, et al. Correctors of Protein Trafficking Defects Identified by a Novel High-Throughput Screening Assay. *ChemBioChem*. 2007 Jun 18;8(9):1012–20.
- Castellani C, Assael BM. Cystic fibrosis: a clinical view. *Cell Mol Life Sci*. 2017 Jan;74(1):129–40.
- Cheng KC, Lin RJ, Cheng JY, Wang SH, Yu JC, Wu JC, et al. FAM129B, an antioxidative protein, reduces chemosensitivity by competing with Nrf2 for Keap1 binding. *EBioMedicine*. 2019 Jul;45:25–38.
- Cheng SH, Gregory RJ, Marshall J, Paul S, Souza DW, White GA, et al. Defective intracellular transport and processing of CFTR is the molecular basis of most cystic fibrosis. *Cell*. 1990 Nov;63(4):827–34.
- Chisti Y, Moo-Young M. Disruption of microbial cells for intracellular products. *Enzyme and Microbial Technology*. 1986 Apr;8(4):194–204.
- Cholon DM, Quinney NL, Fulcher ML, Esther CR, Das J, Dokholyan NV, et al. Potentiator ivacaftor abrogates pharmacological correction of $\Delta F508$ CFTR in cystic fibrosis. *Sci Transl Med [Internet]*. 2014 Jul 23 [cited 2022 Jun 21];6(246). Available from: <https://www.science.org/doi/10.1126/scitranslmed.3008680>
- Clancy JP, Rowe SM, Accurso FJ, Aitken ML, Amin RS, Ashlock MA, et al. Results of a phase IIa study of VX-809, an investigational CFTR corrector compound, in subjects with cystic fibrosis homozygous for the *F508del-CFTR* mutation. *Thorax*. 2012 Jan;67(1):12–8.
- Clarke LA, Sousa L, Barreto C, Amaral MD. Changes in transcriptome of native nasal epithelium expressing F508del-CFTR and intersecting data from comparable studies. *Respir Res*. 2013;14(1):38.
- Collins F. Cystic fibrosis: molecular biology and therapeutic implications. *Science*. 1992 May 8;256(5058):774–9.
- Cooper PJ, Robertson CF, Hudson IL, Phelan PD. Variability of pulmonary function tests in cystic fibrosis. *Pediatr Pulmonol*. 1990;8(1):16–22.

Corey M, Edwards L, Levison H, Knowles M. Longitudinal analysis of pulmonary function decline in patients with cystic fibrosis. *The Journal of Pediatrics*. 1997 Dec;131(6):809–14.

Corey M, Edwards L, Levison H, Knowles M. Longitudinal analysis of pulmonary function decline in patients with cystic fibrosis. *The Journal of Pediatrics*. 1997 Dec;131(6):809–14.

Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol*. 2008 Dec;26(12):1367–72.

Crenesse D, Berlioz M, Bourrier T, Albertini M. Spirometry in children aged 3 to 5 years: Reliability of forced expiratory maneuvers. *Pediatr Pulmonol*. 2001 Jul;32(1):56–61.

Csanády L, Vergani P, Gadsby DC. Structure, Gating, and Regulation of the CFTR Anion Channel. *Physiological Reviews*. 2019 Jan 1;99(1):707–38.

Csősz É, Emri G, Kalló G, Tsaprailis G, Tózsér J. Highly abundant defense proteins in human sweat as revealed by targeted proteomics and label-free quantification mass spectrometry. *J Eur Acad Dermatol Venereol*. 2015 Oct;29(10):2024–31.

Dallacasagrande V, Hajjar KA. Annexin A2 in Inflammation and Host Defense. *Cells*. 2020 Jun 19;9(6):1499.

Davies JC, Alton EFW, Bush A. Cystic fibrosis. *BMJ*. 2007 Dec 15;335(7632):1255–9.

Davies JC, Cunningham S, Harris WT, Lapey A, Regelman WE, Sawicki GS, et al. Safety, pharmacokinetics, and pharmacodynamics of ivacaftor in patients aged 2–5 years with cystic fibrosis and a CFTR gating mutation (KIWI): an open-label, single-arm study. *The Lancet Respiratory Medicine*. 2016 Feb;4(2):107–15.

Davies JC, Moskowitz SM, Brown C, Horsley A, Mall MA, McKone EF, et al. VX-659–Tezacaftor–Ivacaftor in Patients with Cystic Fibrosis and One or Two Phe508del Alleles. *N Engl J Med*. 2018 Oct 25;379(17):1599–611.

Davies JC, Sermet-Gaudelus I, Naehrlich L, Harris RS, Campbell D, Ahluwalia N, et al. A phase 3, double-blind, parallel-group study to evaluate the efficacy and safety of tezacaftor in combination with ivacaftor in participants 6 through 11 years of age with cystic fibrosis homozygous for F508del or heterozygous for the F508del-CFTR mutation and a residual function mutation. *Journal of Cystic Fibrosis*. 2021 Jan;20(1):68–77.

Davies JC, Van de Steen O, van Koningsbruggen-Rietschel S, Drevinek P, Derichs N, McKone EF, et al. GLPG1837, a CFTR potentiator, in p.Gly551Asp (G551D)-CF patients: An open-label, single-arm, phase 2a study (SAPHIRA1). *Journal of Cystic Fibrosis*. 2019 Sep;18(5):693–9.

Davies JC, Wainwright CE, Canny GJ, Chilvers MA, Howenstine MS, Munck A, et al. Efficacy and Safety of Ivacaftor in Patients Aged 6 to 11 Years with Cystic Fibrosis with a *G551D* Mutation. *Am J Respir Crit Care Med*. 2013 Jun;187(11):1219–25.

Davies JC, Wainwright CE, Sawicki GS, Higgins MN, Campbell D, Harris C, et al. Ivacaftor in Infants Aged 4 to <12 Months with Cystic Fibrosis and a Gating Mutation. Results of a Two-Part Phase 3 Clinical Trial. *Am J Respir Crit Care Med*. 2021 Mar 1;203(5):585–93.

Davis SD, Rosenfeld M, Kerby GS, Brumback L, Kloster MH, Acton JD, et al. Multicenter Evaluation of Infant Lung Function Tests as Cystic Fibrosis Clinical Trial Endpoints. *Am J Respir Crit Care Med*. 2010 Dec;182(11):1387–97.

De Boeck K. Cystic fibrosis in the year 2020: A disease with a new face. *Acta Paediatr*. 2020 May;109(5):893–9.

de Jong PA. Progression of lung disease on computed tomography and pulmonary function tests in children and adults with cystic fibrosis. *Thorax*. 2005 Oct 21;61(1):80–5.

de Jong PA, Nakano Y, Lequin MH, Mayo JR, Woods R, Pare PD, et al. Progressive damage on high resolution computed tomography despite stable lung function in cystic fibrosis. *European Respiratory Journal*. 2004 Jan 1;23(1):93–7.

de Jong PA, Ottink MD, Robben SGF, Lequin MH, Hop WCJ, Hendriks JJE, et al. Pulmonary Disease Assessment in Cystic Fibrosis: Comparison of CT Scoring Systems and Value of Bronchial and Arterial Dimension Measurements. *Radiology*. 2004 May;231(2):434–9.

de Nooijer RA, Nobel JM, Arets HGM, Bot AG, van Berkhout FT, de Rijke YB, et al. Assessment of CFTR function in homozygous R117H-7T subjects. *Journal of Cystic Fibrosis*. 2011 Sep;10(5):326–32.

de Souza DAS, Faucz FR, Pereira-Ferrari L, Sotomaior VS, Raskin S. Congenital bilateral absence of the vas deferens as an atypical form of cystic fibrosis: reproductive implications and genetic counseling. *Andrology*. 2018 Jan;6(1):127–35.

Debray D, Kelly D, Houwen R, Strandvik B, Colombo C. Best practice guidance for the diagnosis and management of cystic fibrosis-associated liver disease. *Journal of Cystic Fibrosis*. 2011 Jun;10:S29–36.

Demeyer S, De Boeck K, Witters P, Cosaert K. Beyond pancreatic insufficiency and liver disease in cystic fibrosis. *Eur J Pediatr*. 2016 Jul;175(7):881–94.

DI SANT'AGNESE. Abnormal electrolyte composition of sweat in cystic fibrosis of the pancreas; clinical significance and relationship to the disease. *Pediatrics*. 1953 Nov 12;

Dill EJ, Dawson R, Sellers DE, Robinson WM, Sawicki GS. Longitudinal Trends in Health-Related Quality of Life in Adults With Cystic Fibrosis. *Chest*. 2013 Sep;144(3):981–9.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013 Jan;29(1):15–21.

Donaldson SH, Pilewski JM, Griesse M, Cooke J, Viswanathan L, Tullis E, et al. Tezacaftor/Ivacaftor in Subjects with Cystic Fibrosis and *F508del* / *F508del*-*CFTR* or *F508del* / *G551D*-*CFTR*. *Am J Respir Crit Care Med*. 2018 Jan 15;197(2):214–24.

Duchen MR. Mitochondria in health and disease: perspectives on a new mitochondrial biology. *Molecular Aspects of Medicine*. 2004 Aug;25(4):365–451.

Dugueperoux I. The *CFTR* 3849+10kbC->T and 2789+5G->A alleles are associated with a mild CF phenotype. *European Respiratory Journal*. 2005 Mar 1;25(3):468–73.

Dukovski D, Villella A, Bastos C, King R, Finley D, Kelly JW, et al. Amplifiers co-translationally enhance *CFTR* biosynthesis via *PCBP1*-mediated regulation of *CFTR* mRNA. *Journal of Cystic Fibrosis*. 2020 Sep;19(5):733–41.

Durmowicz AG, Witzmann KA, Rosebraugh CJ, Chowdhury BA. Change in Sweat Chloride as a Clinical End Point in Cystic Fibrosis Clinical Trials. *Chest*. 2013 Jan;143(1):14–8.

Durmowicz AG, Witzmann KA, Rosebraugh CJ, Chowdhury BA. Change in Sweat Chloride as a Clinical End Point in Cystic Fibrosis Clinical Trials. *Chest*. 2013 Jan;143(1):14–8.

Efron B. Missing Data, Imputation, and the Bootstrap. *Journal of the American Statistical Association*. 1994 Jun;89(426):463–75.

Elborn JS. Cystic fibrosis. *The Lancet*. 2016 Nov;388(10059):2519–31.

Elborn JS, Ramsey BW, Boyle MP, Konstan MW, Huang X, Marigowda G, et al. Efficacy and safety of lumacaftor/ivacaftor combination therapy in patients with cystic fibrosis homozygous for Phe508del *CFTR* by pulmonary function subgroup: a pooled analysis. *The Lancet Respiratory Medicine*. 2016 Aug;4(8):617–26.

Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods*. 2007 Mar;4(3):207–14.

ellis G, Dehillotte C, Lemonnier L. Registre Français De La Mucoviscidose. Bilan Des Données 2015. Vaincre la Mucoviscidose et Institut national d'études démographiques. 2015;

EMA. Clinical development of medicinal products for the treatment of cystic fibrosis. 2009;

Estabrooks SK, Brodsky JL. Ubiquitination of disease-causing CFTR variants in a microsome-based assay. *Analytical Biochemistry*. 2020 Sep;604:113829.

Esther CR, Muhlebach MS, Ehre C, Hill DB, Wolfgang MC, Kesimer M, et al. Mucus accumulation in the lungs precedes structural changes and infection in children with cystic fibrosis. *Sci Transl Med*. 2019 Apr 3;11(486):eaav3488.

Fan J, Salathia N, Liu R, Kaeser GE, Yung YC, Herman JL, et al. Characterizing transcriptional heterogeneity through pathway and gene set overdispersion analysis. *Nat Methods*. 2016 Mar;13(3):241–4.

Farinha CM, King-Underwood J, Sousa M, Correia AR, Henriques BJ, Roxo-Rosa M, et al. Revertants, Low Temperature, and Correctors Reveal the Mechanism of F508del-CFTR Rescue by VX-809 and Suggest Multiple Agents for Full Correction. *Chemistry & Biology*. 2013 Jul;20(7):943–55.

Farinha CM, Nogueira P, Mendes F, Penque D, Amaral MD. The human DnaJ homologue (Hdj)-1/heat-shock protein (Hsp) 40 co-chaperone is required for the in vivo stabilization of the cystic fibrosis transmembrane conductance regulator by Hsp70. *Biochemical Journal*. 2002 Sep 15;366(3):797–806.

Farrell PM, Rosenstein BJ, White TB, Accurso FJ, Castellani C, Cutting GR, et al. Guidelines for Diagnosis of Cystic Fibrosis in Newborns through Older Adults: Cystic Fibrosis Foundation Consensus Report. *The Journal of Pediatrics*. 2008 Aug;153(2):S4–14.

Favia M., Atlante A. Mitochondria and cystic fibrosis transmembrane conductance regulator dialogue: Some news. *J Rare Dis Res Treat*. 2016;

FDA. Guidance for industry: Patient-reported outcome measures: use in medical product development to support labeling claims. 2009;

Feist P, Hummon A. Proteomic Challenges: Sample Preparation Techniques for Microgram-Quantity Protein Analysis from Biological Samples. *IJMS*. 2015 Feb 5;16(2):3537–63.

Fidler MC, Beusmans J, Panorchan P, Van Goor F. Correlation of sweat chloride and percent predicted FEV1 in cystic fibrosis patients treated with ivacaftor. *Journal of Cystic Fibrosis*. 2017 Jan;16(1):41–4.

Flume PA, Liou TG, Borowitz DS, Li H, Yen K, Ordoñez CL, et al. Ivacaftor in Subjects With Cystic Fibrosis Who Are Homozygous for the F508del-CFTR Mutation. *Chest*. 2012 Sep;142(3):718–24.

Frischer T, Myung JK, Maurer G, Eichler I, Szepfalusi Z, Lubec G. Possible dysregulation of chaperon and metabolic proteins in cystic fibrosis bronchial tissue. *Proteomics*. 2006 Jun;6(11):3381–8.

Frost FJ, Nazareth DS, Charman SC, Winstanley C, Walshaw MJ. Ivacaftor Is Associated with Reduced Lung Infection by Key Cystic Fibrosis Pathogens. A Cohort Study Using National Registry Data. *Annals ATS*. 2019 Nov;16(11):1375–82.

Frost F, Nazareth D, Shaw M, Al-Aloul M. Assessing the validity and applicability of the French 3-year prognostic score in the UK cystic fibrosis population – a national cohort study. *Transpl Int*. 2021 Mar;34(3):436–44.

Gadsby DC, Vergani P, Csanády L. The ABC protein turned chloride channel whose failure causes cystic fibrosis. *Nature*. 2006 Mar;440(7083):477–83.

Galiotta LVJ, Jayaraman S, Verkman AS. Cell-based assay for high-throughput quantitative screening of CFTR chloride transport agonists. *American Journal of Physiology-Cell Physiology*. 2001 Nov 1;281(5):C1734–42.

Garcia-Alonso L, Holland CH, Ibrahim MM, Turei D, Saez-Rodriguez J. Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res*. 2019 Aug;29(8):1363–75.

Garçon-Michel N, Roguedas-Contios AM, Rault G, Le Bihan J, Ramel S, Revert K, et al. Frequency of aquagenic palmoplantar keratoderma in cystic fibrosis: a new sign of cystic fibrosis?: Aquagenic palmoplantar keratoderma in cystic fibrosis. *British Journal of Dermatology*. 2010 Jul;163(1):162–6.

Geitani R, Moubareck CA, Xu Z, Karam Sarkis D, Touqui L. Expression and Roles of Antimicrobial Peptides in Innate Defense of Airway Mucosa: Potential Implication in Cystic Fibrosis. *Front Immunol*. 2020 Jun 30;11:1198.

Giansanti P, Tsiatsiani L, Low TY, Heck AJR. Six alternative proteases for mass spectrometry-based proteomics beyond trypsin. *Nat Protoc*. 2016 May;11(5):993–1006.

Gorban AN, Kégl B, Wunsch DC, Zinovyev AY, editors. *Principal Manifolds for Data Visualization and Dimension Reduction* [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 2008 [cited 2022 Oct 29]. (Barth TJ, Griebel M, Keyes DE, Nieminen RM, Roose D, Schlick T, editors. *Lecture Notes in Computational Science and Engineering*; vol. 58). Available from: <http://link.springer.com/10.1007/978-3-540-73750-6>

Guimbellot JS, Baines A, Paynter A, Heltshe SL, VanDalfsen J, Jain M, et al. Long term clinical effectiveness of ivacaftor in people with the G551D CFTR mutation. *Journal of Cystic Fibrosis*. 2021 Mar;20(2):213–9.

Gustafsson PM, Aurora P, Lindblad A. Evaluation of ventilation maldistribution as an early indicator of lung disease in children with cystic fibrosis. *European Respiratory Journal*. 2003 Dec 1;22(6):972–9.

H. C. Role of actin filament organization in CFTR activation. *Pflgers Archiv European Journal of Physiology*. 2001 Nov 1;443(0):S75–80.

Harshman SW, Pitsch RL, Smith ZK, O'Connor ML, Geier BA, Qualley AV, et al. The proteomic and metabolomic characterization of exercise-induced sweat for human performance monitoring: A pilot investigation. Koomen JM, editor. *PLoS ONE*. 2018 Nov 1;13(11):e0203133.

Hastie T, Friedman J, Tibshirani R. *The Elements of Statistical Learning* [Internet]. New York, NY: Springer New York; 2001 [cited 2022 Oct 29]. (Springer Series in Statistics). Available from: <http://link.springer.com/10.1007/978-0-387-21606-5>

Hawkins RD, Hon GC, Ren B. Next-generation genomics: an integrative approach. *Nat Rev Genet*. 2010 Jul;11(7):476–86.

Heijerman HGM, McKone EF, Downey DG, Van Braeckel E, Rowe SM, Tullis E, et al. Efficacy and safety of the elexacaftor plus tezacaftor plus ivacaftor combination regimen in people with cystic fibrosis homozygous for the F508del mutation: a double-blind, randomised, phase 3 trial. *The Lancet*. 2019 Nov;394(10212):1940–8.

Heltshe SL, Mayer-Hamblett N, Burns JL, Khan U, Baines A, Ramsey BW, et al. *Pseudomonas aeruginosa* in Cystic Fibrosis Patients With G551D-CFTR Treated With Ivacaftor. *Clinical Infectious Diseases*. 2015 Mar 1;60(5):703–12.

Hotelling H. Analysis of a Complex of Statistical Variables Into Principal Components. *Journal of Educational Psychology*. 1933;417–41 and 498–520.

Hubert D, Marguet C, Benichou J, DeSouza C, Payen-Champenois C, Kinnman N, et al. Real-World Long-Term Ivacaftor for Cystic Fibrosis in France: Clinical Effectiveness and Healthcare Resource Utilization. *Pulm Ther*. 2021 Dec;7(2):455–68.

Hubert LJ, Levin JR. A general statistical framework for assessing categorical clustering in free recall. *Psychological Bulletin*. 1976;83(6):1072–80.

Hudson RP, Dawson JE, Chong PA, Yang Z, Millen L, Thomas PJ, et al. Direct Binding of the Corrector VX-809 to Human CFTR NBD1: Evidence of an Allosteric Coupling between the Binding Site and the NBD1:CL4 Interface. *Mol Pharmacol*. 2017 Aug;92(2):124–35.

Ihaka R, Gentleman R. R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics*. 1996 Sep;5(3):299.

International Human Genome Sequencing Consortium, Whitehead Institute for Biomedical Research, Center for Genome Research:, Lander ES, Linton LM, Birren B, Nusbaum C, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001 Feb 15;409(6822):860–921.

Jacobi E, Solomon M, Avolio J, Shaw M, Gonska T, Ratjen F, et al. Aquagenic wrinkling of the palms in cystic fibrosis patients treated with ivacaftor. *Journal of Cystic Fibrosis*. 2022 Mar;21(2):e102–5.

Jensen R, Stanojevic S, Gibney K, Salazar JG, Gustafsson P, Subbarao P, et al. Multiple Breath Nitrogen Washout: A Feasible Alternative to Mass Spectrometry. Hartl D, editor. *PLoS ONE*. 2013 Feb 15;8(2):e56868.

Jin L, Bi Y, Hu C, Qu J, Shen S, Wang X, et al. A comparative study of evaluating missing value imputation methods in label-free proteomics. *Sci Rep*. 2021 Dec;11(1):1760.

Jolliffe IT. Principal Component Analysis.

Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. *Phil Trans R Soc A*. 2016 Apr 13;374(2065):20150202.

Josse J, Pagès J, Husson F. Multiple imputation in principal component analysis. *Adv Data Anal Classif*. 2011 Oct;5(3):231–46.

Judge EP, Dodd JD, Masterson JB, Gallagher CG. Pulmonary Abnormalities on High-Resolution CT Demonstrate More Rapid Decline Than FEV₁ in Adults With Cystic Fibrosis. *Chest*. 2006 Nov;130(5):1424–32.

Kasperek P, Ileninova Z, Zbodakova O, Kanchev I, Benada O, Chalupsky K, et al. KLK5 and KLK7 Ablation Fully Rescues Lethality of Netherton Syndrome-Like Phenotype. Barsh GS, editor. *PLoS Genet*. 2017 Jan 17;13(1):e1006566.

Katchman BA, Zhu M, Blain Christen J, Anderson KS. Eccrine Sweat as a Biofluid for Profiling Immune Biomarkers. *Prot Clin Appl*. 2018 Nov;12(6):1800010.

Kawala CR, Ma X, Sykes J, Stanojevic S, Coriati A, Stephenson AL. Real-world use of ivacaftor in Canada: A retrospective analysis using the Canadian Cystic Fibrosis Registry. *Journal of Cystic Fibrosis*. 2021 Nov;20(6):1040–5.

Keating D, Marigowda G, Burr L, Daines C, Mall MA, McKone EF, et al. VX-445–Tezacaftor–Ivacaftor in Patients with Cystic Fibrosis and One or Two Phe508del Alleles. *N Engl J Med*. 2018 Oct 25;379(17):1612–20.

Kelly-Aubert M, Trudel S, Fritsch J, Nguyen-Khoa T, Baudouin-Legros M, Moriceau S, et al. GSH monoethyl ester rescues mitochondrial defects in cystic fibrosis models. *Human Molecular Genetics*. 2011 Jul 15;20(14):2745–59.

Kerem BS, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, et al. Identification of the Cystic Fibrosis Gene: Genetic Analysis. *Science*. 1989 Sep 8;245(4922):1073–80.

Khan MA, Ali ZS, Sweezey N, Grasemann H, Palaniyar N. Progression of Cystic Fibrosis Lung Disease from Childhood to Adulthood: Neutrophils, Neutrophil Extracellular Trap (NET) Formation, and NET Degradation. *Genes*. 2019 Feb 26;10(3):183.

Khan TZ, Wagener JS, Bost T, Martinez J, Accurso FJ, Riches DWH. Early Pulmonary Inflammation in Infants with Cystic Fibrosis. *Am J Respir Crit Care Med*. 1995 Apr;151(4):1075–82.

Konstan MW, McKone EF, Moss RB, Marigowda G, Tian S, Waltz D, et al. Assessment of safety and efficacy of long-term treatment with combination lumacaftor and ivacaftor therapy in patients with cystic fibrosis homozygous for the F508del-CFTR mutation (PROGRESS): a phase 3, extension study. *The Lancet Respiratory Medicine*. 2017 Feb;5(2):107–18.

Kormann MSD, Dewerth A, Eichner F, Baskaran P, Hector A, Regamey N, et al. Transcriptomic profile of cystic fibrosis patients identifies type I interferon response and ribosomal stalk proteins as potential modifiers of disease severity. Chotirmall SH, editor. *PLoS ONE*. 2017 Aug 28;12(8):e0183526.

Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov MN, Sergushichev A. Fast gene set enrichment analysis [Internet]. *Bioinformatics*; 2016 Jun [cited 2022 May 3]. Available from: <http://biorxiv.org/lookup/doi/10.1101/060012>

Kotloff RM, FitzSimmons SC, Fiel SB. Fertility and pregnancy in patients with cystic fibrosis. *Clin Chest Med*. 1992 Dec;13(4):623–35.

Krämer A, Green J, Pollard J, Tugendreich S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics*. 2014 Feb 15;30(4):523–30.

Kwak EJ, Hong JY, Kim MN, Kim SY, Kim SH, Park CO, et al. Chitinase 3-like 1 drives allergic skin inflammation via Th2 immunity and M2 macrophage activation. *Clin Exp Allergy*. 2019 Nov;49(11):1464–74.

Laselva O, Molinski S, Casavola V, Bear CE. Correctors of the Major Cystic Fibrosis Mutant Interact through Membrane-Spanning Domains. *Mol Pharmacol*. 2018 Jun;93(6):612–8.

Leonardi S, Parisi GF, Capizzi A, Manti S, Cuppari C, Scuderi MG, et al. YKL-40 as marker of severe lung disease in cystic fibrosis patients. *Journal of Cystic Fibrosis*. 2016 Sep;15(5):583–6.

Levine DM, Haynor DR, Castle JC, Stepaniants SB, Pellegrini M, Mao M, et al. [No title found]. *Genome Biol*. 2006;7(10):R93.

Lewis HA, Buchanan SG, Burley SK, Connors K, Dickey M, Dorwart M, et al. Structure of nucleotide-binding domain 1 of the cystic fibrosis transmembrane conductance regulator. *EMBO J*. 2004 Jan 28;23(2):282–93.

- Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Systems*. 2015 Dec;1(6):417–25.
- Ling KM, Garratt LW, Gill EE, Lee AHY, Agudelo-Romero P, Sutanto EN, et al. Rhinovirus Infection Drives Complex Host Airway Molecular Responses in Children With Cystic Fibrosis. *Front Immunol*. 2020 Jul 16;11:1327.
- Liou TG, Kanner RE. Spirometry. *Clinic Rev Allerg Immunol*. 2009 Dec;37(3):137–52.
- Liu M, Dongre A. Proper imputation of missing values in proteomics datasets for differential expression analysis. *Briefings in Bioinformatics*. 2021 May 20;22(3):bbaa112.
- liu qiangni. The CFTR Corrector, VX-809 (Lumacaftor), Rescues ABCA4 Trafficking Mutants: a Potential Treatment for Stargardt Disease. *Cell Physiol Biochem*. 2019 Aug 13;53(2):400–12.
- Lopes-Pacheco M. CFTR Modulators: Shedding Light on Precision Medicine for Cystic Fibrosis. *Front Pharmacol* [Internet]. 2016 Sep 5 [cited 2022 Jun 15];7. Available from: <http://journal.frontiersin.org/Article/10.3389/fphar.2016.00275/abstract>
- Lum S, Gustafsson P, Ljungberg H, Hulskamp G, Bush A, Carr SB, et al. Early detection of cystic fibrosis lung disease: multiple-breath washout versus raised volume tests. *Thorax*. 2007 Jan 24;62(4):341–7.
- Lum S, Stocks J, Stanojevic S, Wade A, Robinson P, Gustafsson P, et al. Age and height dependence of lung clearance index and functional residual capacity. *Eur Respir J*. 2013 Jun;41(6):1371–7.
- Lund-Palau H, Turnbull AR, Bush A, Bardin E, Cameron L, Soren O, et al. *Pseudomonas aeruginosa* infection in cystic fibrosis: pathophysiological mechanisms and therapeutic approaches. *Expert Review of Respiratory Medicine*. 2016 Jun 2;10(6):685–97.
- M R Picciotto 1 , J A Cohn, G Bertuzzi, P Greengard, A C Nairn. Phosphorylation of the cystic fibrosis transmembrane conductance regulator. *JBiol.chem*. 1992;
- Mailloux RJ. Mitochondrial Antioxidants and the Maintenance of Cellular Hydrogen Peroxide Levels. *Oxidative Medicine and Cellular Longevity*. 2018 Jul 2;2018:1–10.
- Mailloux RJ, Harper ME. Uncoupling proteins and the control of mitochondrial reactive oxygen species production. *Free Radical Biology and Medicine*. 2011 Sep;51(6):1106–15.
- Marentis TC, Kusler B, Yaralioglu GG, Liu S, Hægström EO, Khuri-Yakub BT. Microfluidic sonicator for real-time disruption of eukaryotic cells and bacterial spores for DNA analysis. *Ultrasound in Medicine & Biology*. 2005 Sep;31(9):1265–77.

Marostica PJC, Weist AD, Eigen H, Angelicchio C, Christoph K, Savage J, et al. Spirometry in 3- to 6-Year-Old Children with Cystic Fibrosis. *Am J Respir Crit Care Med*. 2002 Jul;166(1):67–71.

Marson FAL, Bertuzzo CS, Ribeiro JD. Classification of CFTR mutation classes. *The Lancet Respiratory Medicine*. 2016 Aug;4(8):e37–8.

Martignetti L, Calzone L, Bonnet E, Barillot E, Zinovyev A. ROMA: Representation and Quantification of Module Activity from Target Expression Data. *Front Genet* [Internet]. 2016 Feb 19 [cited 2021 Nov 24];7.

Masson A, Schneider-Futschik EK, Baatallah N, Nguyen-Khoa T, Girodon E, Hatton A, et al. Predictive factors for lumacaftor/ivacaftor clinical response. *Journal of Cystic Fibrosis*. 2019 May;18(3):368–74.

Matos AM, Pinto FR, Barros P, Amaral MD, Pepperkok R, Matos P. Inhibition of calpain 1 restores plasma membrane stability to pharmacologically rescued Phe508del-CFTR variant. *Journal of Biological Chemistry*. 2019 Sep;294(36):13396–410.

Matsui H, Grubb BR, Tarran R, Randell SH, Gatzky JT, Davis CW, et al. Evidence for Periciliary Liquid Layer Depletion, Not Abnormal Ion Composition, in the Pathogenesis of Cystic Fibrosis Airways Disease. *Cell*. 1998 Dec;95(7):1005–15.

McCallum TJ, Milunsky JM, Cunningham DL, Harris DH, Maher TA, Oates RD. Fertility in Men With Cystic Fibrosis. *Chest*. 2000 Oct;118(4):1059–62.

McCarthy C, Dimitrov BD, Meurling IJ, Gunaratnam C, McElvaney NG. The CF-ABLE Score. *Chest*. 2013 May;143(5):1358–64.

McColley SA, Konstan MW, Ramsey BW, Stuart Elborn J, Boyle MP, Wainwright CE, et al. Lumacaftor/Ivacaftor reduces pulmonary exacerbations in patients irrespective of initial changes in FEV1. *Journal of Cystic Fibrosis*. 2019 Jan;18(1):94–101.

Meacham GC. The Hdj-2/Hsc70 chaperone pair facilitates early steps in CFTR biogenesis. *The EMBO Journal*. 1999 Mar 15;18(6):1492–505.

Mehta G, Macek M, Mehta A. Cystic fibrosis across Europe: EuroCareCF analysis of demographic data from 35 countries. *Journal of Cystic Fibrosis*. 2010 Dec;9:S5–21.

Middleton PG, Mall MA, Dřevínek P, Lands LC, McKone EF, Polineni D, et al. Elexacaftor–Tezacaftor–Ivacaftor for Cystic Fibrosis with a Single Phe508del Allele. *N Engl J Med*. 2019 Nov 7;381(19):1809–19.

Milla CE, Ratjen F, Marigowda G, Liu F, Waltz D, Rosenfeld M. Lumacaftor/Ivacaftor in Patients Aged 6–11 Years with Cystic Fibrosis and Homozygous for *F508del-CFTR*. *Am J Respir Crit Care Med*. 2017 Apr;195(7):912–20.

Miller MR. Standardisation of spirometry. *European Respiratory Journal*. 2005 Aug 1;26(2):319–38.

Mitchell P. Coupling of Phosphorylation to Electron and Hydrogen Transfer by a Chemi-Osmotic type of Mechanism. *Nature*. 1961 Jul;191(4784):144–8.

Mitchell RM, Jones AM, Stocking K, Foden P, Barry PJ. Longitudinal effects of ivacaftor and medicine possession ratio in people with the *Gly551Asp* mutation: a 5-year study. *Thorax*. 2021 Sep;76(9):874–9.

Moyer J, Wilson D, Finkelshtein I, Wong B, Potts R. Correlation Between Sweat Glucose and Blood Glucose in Subjects with Diabetes. *Diabetes Technology & Therapeutics*. 2012 May;14(5):398–402.

Murakami M, Ohtake T, Dorschner RA, Gallo RL, Schitteck B, Garbe C. Cathelicidin Anti-Microbial Peptide Expression in Sweat, an Innate Defense System for the Skin. *Journal of Investigative Dermatology*. 2002 Nov;119(5):1090–5.

Murphy MP. How mitochondria produce reactive oxygen species. *Biochemical Journal*. 2009 Jan 1;417(1):1–13.

Murphy MP. Mitochondria--a neglected drug target. *Curr Opin Investig Drugs*. 2009 Oct;10(10):1022–4.

Mutyam V, Libby EF, Peng N, Hadjiliadis D, Bonk M, Solomon GM, et al. Therapeutic benefit observed with the CFTR potentiator, ivacaftor, in a CF patient homozygous for the W1282X CFTR nonsense mutation. *Journal of Cystic Fibrosis*. 2017 Jan;16(1):24–9.

Navarro P, Vázquez J. A Refined Method To Calculate False Discovery Rates for Peptide Identification Using Decoy Databases. *J Proteome Res*. 2009 Apr 3;8(4):1792–6.

Neupert W, Herrmann JM. Translocation of Proteins into Mitochondria. *Annu Rev Biochem*. 2007 Jun 7;76(1):723–49.

Nguyen-Khoa T, Hatton A, Drummond D, Aoust L, Schlatter J, Martin C, et al. Reclassifying inconclusive diagnosis for cystic fibrosis with new generation sweat test. *Eur Respir J*. 2022 Aug;60(2):2200209.

Nkam L, Lambert J, Latouche A, Bellis G, Burgel Pr, Hocine MN. A 3-year prognostic score for adults with cystic fibrosis. *Journal of Cystic Fibrosis*. 2017 Nov;16(6):702–8.

Norez C, Noel S, Wilke M, Bijvelds M, Jorna H, Melin P, et al. Rescue of functional F508del-CFTR channels in cystic fibrosis epithelial cells by the α -glucosidase inhibitor miglustat. *FEBS Letters*. 2006 Apr 3;580(8):2081–6.

O'Sullivan BP, Baker D, Leung KG, Reed G, Baker SS, Borowitz D. Evolution of Pancreatic Function during the First Year in Infants with Cystic Fibrosis. *The Journal of Pediatrics*. 2013 Apr;162(4):808-812.e1.

O'Connell JD, Paulo JA, O'Brien JJ, Gygi SP. Proteome-Wide Evaluation of Two Common Protein Quantification Methods. *J Proteome Res*. 2018 May 4;17(5):1934-42.

Ogilvie V, Passmore M, Hyndman L, Jones L, Stevenson B, Wilson A, et al. Differential global gene expression in cystic fibrosis nasal and bronchial epithelium. *Genomics*. 2011 Nov;98(5):327-36.

Okiyoneda T, Niibori A, Harada K, Kohno T, Michalak M, Duszyk M, et al. Role of calnexin in the ER quality control and productive folding of CFTR; differential effect of calnexin knockout on wild-type and $\Delta F508$ CFTR. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*. 2008 Sep;1783(9):1585-94.

Okuda K, Dang H, Kobayashi Y, Carraro G, Nakano S, Chen G, et al. Secretory Cells Dominate Airway CFTR Expression and Function in Human Airway Superficial Epithelia. *Am J Respir Crit Care Med*. 2021 May 15;203(10):1275-89.

Olesen HV, Drevinek P, Gulmans VA, Hatziaorou E, Jung A, Mei-Zahav M, et al. Cystic fibrosis related diabetes in Europe: Prevalence, risk factors and outcome; Olesen et al. *Journal of Cystic Fibrosis*. 2020 Mar;19(2):321-7.

Palmer CNA, Irvine AD, Terron-Kwiatkowski A, Zhao Y, Liao H, Lee SP, et al. Common loss-of-function variants of the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis. *Nat Genet*. 2006 Apr;38(4):441-6.

Pankow S, Bamberger C, Calzolari D, Martínez-Bartolomé S, Lavallée-Adam M, Balch WE, et al. $\Delta F508$ CFTR interactome remodelling promotes rescue of cystic fibrosis. *Nature*. 2015 Dec 24;528(7583):510-6.

Pankow S, Bamberger C, Yates JR. A posttranslational modification code for CFTR maturation is altered in cystic fibrosis. *Sci Signal*. 2019 Jan;12(562):eaan7984.

Park JH, Park GT, Cho IH, Sim SM, Yang JM, Lee DY. An antimicrobial protein, lactoferrin exists in the sweat: proteomic analysis of sweat: Letter to the Editor. *Experimental Dermatology*. 2011 Apr;20(4):369-71.

Pearson K. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*. 1901;559-572.

Pedemonte N. Small-molecule correctors of defective F508-CFTR cellular processing identified by high-throughput screening. *Journal of Clinical Investigation*. 2005 Aug 25;115(9):2564-71.

- Pedemonte N, Tomati V, Sondo E, Galietta LJ. Influence of cell background on pharmacological rescue of mutant CFTR. *American Journal of Physiology-Cell Physiology*. 2010 Apr;298(4):C866–74.
- Polineni D, Dang H, Gallins PJ, Jones LC, Pace RG, Stonebraker JR, et al. Airway Mucosal Host Defense Is Key to Genomic Regulation of Cystic Fibrosis Lung Disease Severity. *Am J Respir Crit Care Med*. 2018 Jan 1;197(1):79–93.
- Pollitt RJ. Neonatal screening for cystic fibrosis. *Seminars in Neonatology*. 1998 Feb;3(1):9–15.
- Pranke IM, Sermet-Gaudelus I. Biosynthesis of cystic fibrosis transmembrane conductance regulator. *The International Journal of Biochemistry & Cell Biology*. 2014 Jul;52:26–38.
- Puglia M, Landi C, Gagliardi A, Breslin L, Armini A, Brunetti J, et al. The proteome speciation of an immortalized cystic fibrosis cell line: New perspectives on the pathophysiology of the disease. *Journal of Proteomics*. 2018 Jan;170:28–42.
- Putman MS, Anabtawi A, Le T, Tangpricha V, Sermet-Gaudelus I. Cystic fibrosis bone disease treatment: Current knowledge and future directions. *Journal of Cystic Fibrosis*. 2019 Oct;18:S56–65.
- Quinton PM. Chloride impermeability in cystic fibrosis. *Nature*. 1983 Feb;301(5899):421–2.
- Quittner A, Suthoff E, Rendas-Baum R, Bayliss MS, Sermet-Gaudelus I, Castiglione B, et al. Effect of ivacaftor treatment in patients with cystic fibrosis and the G551D-CFTR mutation: patient-reported outcomes in the STRIVE randomized, controlled trial. *Health Qual Life Outcomes*. 2015 Dec;13(1):93.
- Raiszadeh MM, Ross MM, Russo PS, Schaepper MA, Zhou W, Deng J, et al. Proteomic Analysis of Eccrine Sweat: Implications for the Discovery of Schizophrenia Biomarker Proteins. *J Proteome Res*. 2012 Apr 6;11(4):2127–39.
- Ramos KJ, Wai TH, Sykes J, Ma X, Stephenson AL, Jennerich AL, et al. Validation of the French 3-year prognostic score for death or lung transplant in the United States cystic fibrosis population. *Journal of Cystic Fibrosis*. 2022 May;21(3):471–4.
- Ramos-Rodriguez RR, Cuevas-Diaz-Duran R, Falciani F, Tamez-Peña JG, Trevino V. COMPADRE: an R and web resource for pathway activity analysis by component decompositions. *Bioinformatics*. 2012 Oct 15;28(20):2701–2.
- Ramsey BW, Davies J, McElvaney NG, Tullis E, Bell SC, Dřevínek P, et al. A CFTR Potentiator in Patients with Cystic Fibrosis and the *G551D* Mutation. *N Engl J Med*. 2011 Nov 3;365(18):1663–72.

Ramsey KA, Rosenow T, Turkovic L, Skoric B, Banton G, Adams AM, et al. Lung Clearance Index and Structural Lung Disease on Computed Tomography in Early Cystic Fibrosis. *Am J Respir Crit Care Med*. 2016 Jan;193(1):60–7.

Ramsey KA, Schultz A, Stick SM. Biomarkers in Paediatric Cystic Fibrosis Lung Disease. *Paediatric Respiratory Reviews*. 2015 Sep;16(4):213–8.

Ratjen F, Durham T, Navratil T, Schaberg A, Accurso FJ, Wainwright C, et al. Long term effects of denufosol tetrasodium in patients with cystic fibrosis. *Journal of Cystic Fibrosis*. 2012 Dec;11(6):539–49.

Ratjen F, Hug C, Marigowda G, Tian S, Huang X, Stanojevic S, et al. Efficacy and safety of lumacaftor and ivacaftor in patients aged 6–11 years with cystic fibrosis homozygous for F508del-CFTR: a randomised, placebo-controlled phase 3 trial. *The Lancet Respiratory Medicine*. 2017 Jul;5(7):557–67.

Ratjen F, Klingel M, Black P, Powers MR, Grasemann H, Solomon M, et al. Changes in Lung Clearance Index in Preschool-aged Patients with Cystic Fibrosis Treated with Ivacaftor (GOAL): A Clinical Trial. *Am J Respir Crit Care Med*. 2018 Aug 15;198(4):526–8.

Rehman T, Karp PH, Tan P, Goodell BJ, Pezzulo AA, Thurman AL, et al. Inflammatory cytokines TNF- α and IL-17 enhance the efficacy of cystic fibrosis transmembrane conductance regulator modulators. *Journal of Clinical Investigation*. 2021 Aug 16;131(16):e150398.

Ren HY, Grove DE, De La Rosa O, Houck SA, Sopha P, Van Goor F, et al. VX-809 corrects folding defects in cystic fibrosis transmembrane conductance regulator protein through action on membrane-spanning domain 1. *Gilmore R, editor. MBoC*. 2013 Oct;24(19):3016–24.

Ridgeway ME, Lubeck M, Jordens J, Mann M, Park MA. Trapped ion mobility spectrometry: A short review. *International Journal of Mass Spectrometry*. 2018 Feb;425:22–35.

Riordan JR, Rommens JM, Kerem BS, Alon N, Rozmahel R, Grzelczak Z, et al. Identification of the Cystic Fibrosis Gene: Cloning and Characterization of Complementary DNA. *Science*. 1989 Sep 8;245(4922):1066–73.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*. 2015 Apr 20;43(7):e47–e47.

Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010 Jan 1;26(1):139–40.

Robinson PD, Goldman MD, Gustafsson PM. Inert Gas Washout: Theoretical Background and Clinical Utility in Respiratory Disease. *Respiration*. 2009;78(3):339–55.

Roda J, Pinto-Silva C, Silva IAI, Maia C, Almeida S, Ferreira R, et al. New drugs in cystic fibrosis: what has changed in the last decade? *Therapeutic Advances in Chronic Disease*. 2022 Jan;13:204062232210981.

Rommens JM, Iannuzzi MC, Kerem BS, Drumm ML, Melmer G, Dean M, et al. Identification of the Cystic Fibrosis Gene: Chromosome Walking and Jumping. *Science*. 1989 Sep 8;245(4922):1059–65.

Rongioletti F, Tomasini C, Crovato F, Marchesi L. Aquagenic (pseudo) keratoderma: a clinical series with new pathological insights: Aquagenic (pseudo) keratoderma. *British Journal of Dermatology*. 2012 Sep;167(3):575–82.

Rowe SM, Accurso F, Clancy JP. Detection of Cystic Fibrosis Transmembrane Conductance Regulator Activity in Early-Phase Clinical Trials. *Proceedings of the American Thoracic Society*. 2007 Aug 1;4(4):387–98.

Rowe SM. Lumacaftor/Ivacaftor Treatment of Patients with Cystic Fibrosis Heterozygous for F508del-CFTR. *Ann Am Thorac Soc* . 2017;

Rowe SM, Clancy JP, Wilschanski M. Nasal Potential Difference Measurements to Assess CFTR Ion Channel Activity. In: Amaral MD, Kunzelmann K, editors. *Cystic Fibrosis* [Internet]. Totowa, NJ: Humana Press; 2011 [cited 2022 Jun 18]. p. 69–86. (Methods in Molecular Biology; vol. 741). Available from: http://link.springer.com/10.1007/978-1-61779-117-8_6

Roxo-Rosa M, Xu Z, Schmidt A, Neto M, Cai Z, Soares CM, et al. Revertant mutants G550E and 4RK rescue cystic fibrosis mutants in the first nucleotide-binding domain of CFTR by different mechanisms. *Proc Natl Acad Sci USA*. 2006 Nov 21;103(47):17891–6.

Ruffin M, Mercier J, Calmel C, Mézinè J, Bigot J, Sutanto EN, et al. Update on SLC6A14 in lung and gastrointestinal physiology and physiopathology: focus on cystic fibrosis. *Cell Mol Life Sci*. 2020 Sep;77(17):3311–23.

Sagel SD, Khan U, Heltshe SL, Clancy JP, Borowitz D, Gelfond D, et al. Clinical Effectiveness of Lumacaftor/Ivacaftor in Patients with Cystic Fibrosis Homozygous for F508del-CFTR. A Clinical Trial. *Annals ATS*. 2021 Jan;18(1):75–83.

Sagel SD, Khan U, Heltshe SL, Clancy JP, Borowitz D, Gelfond D, et al. Clinical Effectiveness of Lumacaftor/Ivacaftor in Patients with Cystic Fibrosis Homozygous for F508del-CFTR. A Clinical Trial. *Annals ATS*. 2021 Jan;18(1):75–83.

Saint-Criq V, Delpiano L, Casement J, Onuora JC, Lin J, Gray MA. Choice of Differentiation Media Significantly Impacts Cell Lineage and Response to CFTR Modulators in Fully Differentiated Primary Cultures of Cystic Fibrosis Human Airway Epithelial Cells. *Cells*. 2020 Sep 21;9(9):2137.

- Sanders DB, Li Z, Brody AS, Farrell PM. Chest computed tomography scores of severity are associated with future lung disease progression in children with cystic fibrosis. *Am J Respir Crit Care Med*. 2011 Oct 1;184(7):816–21.
- Sanders DB, Li Z, Brody AS. Chest computed tomography predicts the frequency of pulmonary exacerbations in children with cystic fibrosis. *Ann Am Thorac Soc*. 2015 Jan;12(1):64–9.
- Sandilands A, Sutherland C, Irvine AD, McLean WHI. Filaggrin in the frontline: role in skin barrier function and disease. *Journal of Cell Science*. 2009 May 1;122(9):1285–94.
- Santos JD, Canato S, Carvalho AS, Botelho HM, Aloria K, Amaral MD, et al. Folding Status Is Determinant over Traffic-Competence in Defining CFTR Interactors in the Endoplasmic Reticulum. *Cells*. 2019 Apr 14;8(4):353.
- Sathe M, Houwen R. Meconium ileus in Cystic Fibrosis. *Journal of Cystic Fibrosis*. 2017 Nov;16:S32–9.
- Scarpulla RC. Transcriptional Paradigms in Mammalian Mitochondrial Biogenesis and Function. *Physiological Reviews*. 2008 Apr;88(2):611–38.
- Schaefer CF, Anthony K, Krupa S, Buchoff J, Day M, Hannay T, et al. PID: the Pathway Interaction Database. *Nucleic Acids Research*. 2009 Jan;37(suppl_1):D674–9.
- Schitteck B, Hipfel R, Sauer B, Bauer J, Kalbacher H, Stevanovic S, et al. Dermcidin: a novel human antibiotic peptide secreted by sweat glands. *Nat Immunol*. 2001 Dec;2(12):1133–7.
- Schneider EK. Cytochrome P450 3A4 Induction: Lumacaftor versus Ivacaftor Potentially Resulting in Significantly Reduced Plasma Concentration of Ivacaftor. *DML*. 2018 Aug 28;12(1):71–4.
- Seidler U, Singh AK, Cinar A, Chen M, Hillesheim J, Hogema B, et al. The Role of the NHERF Family of PDZ Scaffolding Proteins in the Regulation of Salt and Water Transport: Lessons Learned from Knockout Mice. *Annals of the New York Academy of Sciences*. 2009 May;1165(1):249–60.
- Sermet-Gaudelus I, Munck A, Rota M, Roussey M, Feldmann D. Recommandations françaises pour la réalisation et l'interprétation du test de la sueur dans le cadre du dépistage néonatal de la mucoviscidose. *Archives de Pédiatrie*. 2010 Sep;17(9):1349–58.
- Sermet-Gaudelus I, Nguyen-Khoa T, Hatton A, Hayes K, Pranke I. Sweat Chloride Testing and Nasal Potential Difference (NPD) Are Primary Outcome Parameters in Treatment with Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) Modulators. *JPM*. 2021 Jul 27;11(8):729.
- Sermet-Gaudelus I, Vallée B, Urbin I, Torossi T, Marianovski R, Fajac A, et al. Normal Function of the Cystic Fibrosis Conductance Regulator Protein Can Be Associated with Homozygous $\Delta F508$ Mutation. *Pediatr Res*. 2002 Nov;52(5):628–35.

- Shehadul Islam M, Aryasomayajula A, Selvaganapathy P. A Review on Macroscale and Microscale Cell Lysis Methods. *Micromachines*. 2017 Mar 8;8(3):83.
- Shteinberg M, Haq IJ, Polineni D, Davies JC. Cystic fibrosis. *The Lancet*. 2021 Jun;397(10290):2195–211.
- Shwachman H, Mahmoodian A, Neff RK. The sweat test: Sodium and chloride values. *The Journal of Pediatrics*. 1981 Apr;98(4):576–8.
- Sibila O, Laserna E, Shoemark A, Perea L, Bilton D, Crichton ML, et al. Heterogeneity of treatment response in bronchiectasis clinical trials. *Eur Respir J*. 2021 Oct 21;2100777.
- Singh VK, Schwarzenberg SJ. Pancreatic insufficiency in Cystic Fibrosis. *Journal of Cystic Fibrosis*. 2017 Nov;16:S70–8.
- Sly PD, Brennan S, Gangell C, de Klerk N, Murray C, Mott L, et al. Lung Disease at Diagnosis in Infants with Cystic Fibrosis Detected by Newborn Screening. *Am J Respir Crit Care Med*. 2009 Jul 15;180(2):146–52.
- Sonneveld N, Stanojevic S, Amin R, Aurora P, Davies J, Elborn JS, et al. Lung clearance index in cystic fibrosis subjects treated for pulmonary exacerbations. *Eur Respir J*. 2015 Oct;46(4):1055–64.
- Stafler P, Davies JC, Balfour-Lynn IM, Rosenthal M, Bush A. Bronchoscopy in Cystic Fibrosis Infants Diagnosed by Newborn Screening: Bronchoscopy in NBS Infants With CF. *Pediatr Pulmonol*. 2011 Jul;46(7):696–700.
- Stanojevic S, Bowerman C, Robinson P. Multiple breath washout: measuring early manifestations of lung pathology. *Breathe*. 2021 Sep;17(3):210016.
- Stanojevic S, Ratjen F. Physiologic endpoints for clinical studies for cystic fibrosis. *Journal of Cystic Fibrosis*. 2016 Jul;15(4):416–23.
- Stark, R., Grzelak, M. & Hadfield, J. RNA sequencing: the teenage years. *Nat Rev Genet* **20**, 631–656 (2019)
- Stekhoven DJ, Buhlmann P. MissForest--non-parametric missing value imputation for mixed-type data. *Bioinformatics*. 2012 Jan 1;28(1):112–8.
- Stick SM, Brennan S, Murray C, Douglas T, von Ungern-Sternberg BS, Garratt LW, et al. Bronchiectasis in Infants and Preschool Children Diagnosed with Cystic Fibrosis after Newborn Screening. *The Journal of Pediatrics*. 2009 Nov;155(5):623-628.e1.
- Stick SM, Sly PD. Exciting New Clinical Trials in Cystic Fibrosis: Infants Need Not Apply. *Am J Respir Crit Care Med*. 2011 Jun 15;183(12):1577–8.

Stick S, Tiddens H, Aurora P, Gustafsson P, Ranganathan S, Robinson P, et al. Early intervention studies in infants and preschool children with cystic fibrosis: are we ready? *Eur Respir J*. 2013 Aug;42(2):527–38.

Strandvik B. Fatty acid metabolism in cystic fibrosis. Prostaglandins, Leukotrienes and Essential Fatty Acids. 2010 Sep;83(3):121–9.

Subbarao P, Milla C, Aurora P, Davies JC, Davis SD, Hall GL, et al. Multiple-Breath Washout as a Lung Function Test in Cystic Fibrosis. A Cystic Fibrosis Foundation Workshop Report. *Annals ATS*. 2015 Jun;12(6):932–9.

Takatsuki A. Golgi-acting drugs: inducers and inhibitors of Golgi dispersal as probes to analyze Golgi membrane dynamics. In: *Progress in Biotechnology* [Internet]. Elsevier; 2002 [cited 2022 Jun 18]. p. 35–44. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0921042302800417>

Taylor-Cousar JL, Munck A, McKone EF, van der Ent CK, Moeller A, Simard C, et al. Tezacaftor–Ivacaftor in Patients with Cystic Fibrosis Homozygous for Phe508del. *N Engl J Med*. 2017 Nov 23;377(21):2013–23.

The Human Protein Atlas. <https://www.proteinatlas.org/ENSG000000001626-CFTR/tissue>.

Tiddens HAWM, Donaldson SH, Rosenfeld M, Paré PD. Cystic fibrosis lung disease starts in the small airways: Can we treat it more effectively? *Pediatr Pulmonol*. 2010 Feb;45(2):107–17.

Toledano MB, Mukherjee SK, Howell J, Westaby D, Khan SA, Bilton D, et al. The emerging burden of liver disease in cystic fibrosis patients: A UK nationwide study. Strnad P, editor. *PLoS ONE*. 2019 Apr 4;14(4):e0212779.

Tolland JP, Boyle J, Hall V, McKenna KE, Elborn JS. Aquagenic Wrinkling of the Palms in an Adult Cystic Fibrosis Population. *Dermatology*. 2010;221(4):326–30.

Tomfohr J, Lu J, Kepler TB. Pathway level analysis of gene expression using singular value decomposition. *BMC Bioinformatics*. 2005 Dec;6(1):225.

Tsui LC, Buchwald M, Barker D, Braman JC, Knowlton R, Schumm JW, et al. Cystic Fibrosis Locus Defined by a Genetically Linked Polymorphic DNA Marker. *Science*. 1985 Nov 29;230(4729):1054–7.

Turrens JF, Freeman BA, Levitt JG, Crapo JD. The effect of hyperoxia on superoxide production by lung submitochondrial particles. *Archives of Biochemistry and Biophysics*. 1982 Sep;217(2):401–10.

Tyanova S, Cox J. Perseus: A Bioinformatics Platform for Integrative Analysis of Proteomics Data in Cancer Research. In: von Stechow L, editor. *Cancer Systems Biology* [Internet]. New York, NY:

Springer New York; 2018 [cited 2021 Nov 2]. p. 133–48. (Methods in Molecular Biology; vol. 1711). Available from: http://link.springer.com/10.1007/978-1-4939-7493-1_7

Valdivieso AG, Santa-Coloma TA. CFTR activity and mitochondrial function. *Redox Biology*. 2013;1(1):190–202.

Van Goor F, Hadida S, Grootenhuis PDJ, Burton B, Cao D, Neuberger T, et al. Rescue of CF airway epithelial cell function in vitro by a CFTR potentiator, VX-770. *Proc Natl Acad Sci USA*. 2009 Nov 3;106(44):18825–30.

Van Goor F, Straley KS, Cao D, González J, Hadida S, Hazlewood A, et al. Rescue of $\Delta F508$ -CFTR trafficking and gating in human cystic fibrosis airway primary cultures by small molecules. *American Journal of Physiology-Lung Cellular and Molecular Physiology*. 2006 Jun;290(6):L1117–30.

van Koningsbruggen-Rietschel S, Conrath K, Fischer R, Sutharsan S, Kempa A, Gleiber W, et al. GLPG2737 in lumacaftor/ivacaftor-treated CF subjects homozygous for the F508del mutation: A randomized phase 2A trial (PELICAN). *Journal of Cystic Fibrosis*. 2020 Mar;19(2):292–8.

Vandermarliere E, Mueller M, Martens L. Getting intimate with trypsin, the leading protease in proteomics: TRYPSIN IN PROTEOMICS. *Mass Spec Rev*. 2013 Nov;32(6):453–65.

Varga K, Jurkuvenaite A, Wakefield J, Hong JS, Guimbellot JS, Venglarik CJ, et al. Efficient Intracellular Processing of the Endogenous Cystic Fibrosis Transmembrane Conductance Regulator in Epithelial Cell Lines. *Journal of Biological Chemistry*. 2004 May;279(21):22578–84.

Vasconcellos CA, Allen PG, Wohl ME, Drazen JM, Janmey PA, Stossel TP. Reduction in Viscosity of Cystic Fibrosis Sputum in Vitro by Gelsolin. *Science*. 1994 Feb 18;263(5149):969–71.

Verheggen K, Ræder H, Berven FS, Martens L, Barsnes H, Vaudel M. Anatomy and evolution of database search engines—a central component of mass spectrometry based proteomic workflows. *Mass Spec Rev*. 2020 May;39(3):292–306.

Wagener JS, Rasouliyan L, VanDevanter DR, Pasta DJ, Regelman WE, Morgan WJ, et al. Oral, inhaled, and intravenous antibiotic choice for treating pulmonary exacerbations in cystic fibrosis. *Pediatr Pulmonol*. 2013 Jul;48(7):666–73.

Wainwright CE, Elborn JS, Ramsey BW, Marigowda G, Huang X, Cipolli M, et al. Lumacaftor–Ivacaftor in Patients with Cystic Fibrosis Homozygous for Phe508del *CFTR*. *N Engl J Med*. 2015 Jul 16;373(3):220–31.

Walker S, Flume P, McNamara J, Solomon M, Chilvers M, Chmiel J, et al. A phase 3 study of tezacaftor in combination with ivacaftor in children aged 6 through 11 years with cystic fibrosis. *Journal of Cystic Fibrosis*. 2019 Sep;18(5):708–13.

Wang G, Wu WW, Zhang Z, Masilamani S, Shen RF. Decoy Methods for Assessing False Positives and False Discovery Rates in Shotgun Proteomics. *Anal Chem*. 2009 Jan 1;81(1):146–59.

Wang K, Li M, Hakonarson H. Analysing biological pathways in genome-wide association studies. *Nat Rev Genet*. 2010 Dec;11(12):843–54.

Wang X, Venable J, LaPointe P, Hutt DM, Koulov AV, Coppinger J, et al. Hsp90 Cochaperone Aha1 Downregulation Rescues Misfolding of CFTR in Cystic Fibrosis. *Cell*. 2006 Nov;127(4):803–15.

WARD CL. Intracellular turnover of cystic fibrosis transmembrane conductance regulator. Inefficient processing and rapid degradation of wild-type and mutant proteins. *J Biol Chem* . 1994;

Weidinger S, Illig T, Baurecht H, Irvine AD, Rodriguez E, Diaz-Lacava A, et al. Loss-of-function variations within the filaggrin gene predispose for atopic dermatitis with allergic sensitizations. *Journal of Allergy and Clinical Immunology*. 2006 Jul;118(1):214–9.

Welsh MJ, Smith AE. Molecular mechanisms of CFTR chloride channel dysfunction in cystic fibrosis. *Cell*. 1993 Jul;73(7):1251–4.

Wieczorek S, Combes F, Lazar C, Gai Gianetto Q, Gatto L, Dorffer A, et al. DAPAR & ProStaR: software to perform statistical analyses in quantitative discovery proteomics. *Bioinformatics*. 2017 Jan 1;33(1):135–6.

Wright JM, Merlo CA, Reynolds JB, Zeitlin PL, Garcia JGN, Guggino WB, et al. Respiratory Epithelial Gene Expression in Patients with Mild and Severe Cystic Fibrosis Lung Disease. *Am J Respir Cell Mol Biol*. 2006 Sep;35(3):327–36.

Yamashiro Y, Shimizu T, Oguchi S, Shioya T, Nagata S, Ohtsuka Y. The Estimated Incidence of Cystic Fibrosis in Japan: *Journal of Pediatric Gastroenterology & Nutrition*. 1997 May;24(5):544–7.

Youker RT, Walsh P, Beilharz T, Lithgow T, Brodsky JL. Distinct Roles for the Hsp40 and Hsp90 Molecular Chaperones during Cystic Fibrosis Transmembrane Conductance Regulator Degradation in Yeast. *MBoC*. 2004 Nov;15(11):4787–97.

Yu Y, Prassas I, Muytjens CMJ, Diamandis EP. Proteomic and peptidomic analysis of human sweat with emphasis on proteolysis. *Journal of Proteomics*. 2017 Feb;155:40–8.

Yu YC, Sohma Y, Hwang TC. On the mechanism of gating defects caused by the R117H mutation in cystic fibrosis transmembrane conductance regulator: Gating defects in R117H-CFTR. *J Physiol*. 2016 Jun 15;594(12):3227–44.

Zabner J, Scheetz TE, Almabrazi HG, Casavant TL, Huang J, Keshavjee S, et al. CFTR $\Delta F508$ mutation has minimal effect on the gene expression profile of differentiated human airway epithelia. *American Journal of Physiology-Lung Cellular and Molecular Physiology*. 2005 Oct;289(4):L545–53.

Zemanick ET, Taylor-Cousar JL, Davies J, Gibson RL, Mall MA, McKone EF, et al. A Phase 3 Open-Label Study of Elexacaftor/Tezacaftor/Ivacaftor in Children 6 through 11 Years of Age with Cystic Fibrosis and at Least One *F508del* Allele. *Am J Respir Crit Care Med*. 2021 Jun 15;203(12):1522–32.

Zhang X. Detergents: Friends not foes for high-performance membrane proteomics toward precision medicine. *Proteomics*. 2017 Feb;17(3–4):1600209.

Zhang Y, Fonslow BR, Shan B, Baek MC, Yates JR. Protein Analysis by Shotgun/Bottom-up Proteomics. *Chem Rev*. 2013 Apr 10;113(4):2343–94.

NORD's Rare Disease Database.

Registre Français de la Mucoviscidose 2020.

Walters S, Mehta A. Epidemiology of cystic fibrosis. In: Hodson M, Geddes DM, Bush A, eds. *Cystic fibrosis*, 3rd edn. London: Edward Arnold Ltd, 2007: 21–45. In.

The Raised Volume Rapid Thoracoabdominal Compression Technique: This Workshop Summary was approved by the ATS Board of Directors, December 1999 and The ERS Executive Committee, April 2000. *Am J Respir Crit Care Med*. 2000 May;161(5):1760–2.

Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology & Therapeutics*. 2001 Mar;69(3):89–95.

Lloyd, S. P. (1957). Least squares quantization in PCM. Technical Report RR-5497, Bell Lab, September 1957.

Annexes

ANNEXE 1

Patients characteristics at baseline and at 6 months lumacaftor-ivacaftor using imputed ITT data. Values in n(%) or mean (SD). p values are calculated by paired t-test, or Wilcoxon signed rank test when comparing 6-12 to 12-18 year-old children.

	6-18 years n = 112			6-12 years n = 24			12-18 years old (n = 88)			p 6-12 vs 12-18 yrs	
	M0	M6	p	M0	M6	p	M0	M6	p	M0	M6
Female, n(%)	53 (47)	ND	ND	10 (42)	ND	ND	43 (49)	ND	ND	ND	ND
Age (yrs)	13.5 (2.8)	14.1 (2.9)	ND	9.3 (2.1)	9.8 (2.1)	ND	14.7 (1.7)	15.2 (1.8)	ND	ND	ND
BMI Z-score	-0.6 (0.9)	-0.4 (0.9)	<10e-4	-0.65 (0.9)	-0.4 (1)	0.001	-0.6 (0.9)	-0.4 (0.8)	0.001	0.7 (NS)	0.8 (NS)
ppFEV ₁	87 (23)	89 (21)	0.1 (NS)	100 (20)	101 (20)	0.6 (NS)	83 (22)	86 (20)	0.1 (NS)	<10e-3	0.001
ppFVC	93 (19)	95 (17)	0.02	99 (13)	101 (13)	0.2 (NS)	91 (20)	93 (18)	0.06 (NS)	0.07 (NS)	0.02
ppFEF ₂₅₋₇₅	74 (32)	75 (30)	0.68 (NS)	90 (35)	89 (34)	0.8 (NS)	69 (29)	71 (28)	0.6 (NS)	0.01	0.02
Sweat test (mmol/L)	101 (11)	80 (17)	<10e-14	103 (8)	80 (14)	<10e-4	100 (12)	79 (17)	<10e-11	0.1 (NS)	0.9 (NS)
days of ATB in the previous 6 months	22 (23)	17 (23)	0.01	22 (25)	13 (21)	0.1 (NS)	22 (23)	18 (23)	0.05 (NS)	0.8 (NS)	0.4 (NS)

BMI: Body Mass Index ; ppFEV₁: percent predicted Forced Expiratory Volume in 1s ; ppFVC: percent predicted Forced Vital Capacity ; ppFEF₂₅₋₇₅: percent predicted Forced Expiratory Flow between 25% and 75% of FVC; ATB: antibiotic ; ND: No data to display ; NS: not significant

ANNEXE 2

Correlation coefficients and corresponding p values between the 6 outcome parameters and the 6 Principal Components for patients aged 12 years and older. Percentage of explained variance for each PC.

	PC 1		PC 2		PC 3		PC 4		PC 5		PC 6	
	Corr	p	Corr	p	Corr	p	Corr	p	Corr	p	Corr	p
ΔBMI Z-score	0.09	0.2 (NS)	-0.83	<10e-15	0.16	0.1 (NS)	0.52	<10e-6	-0.02	0.9 (NS)	-0.03	0.8 (NS)
ΔppFEV ₁	0.95	<10e-15	0.07	0.5 (NS)	-0.13	0.2 (NS)	-0.05	0.6 (NS)	-0.09	0.4 (NS)	-0.24	0.02
ΔppFVC	0.85	<10e-15	-0.05	0.7 (NS)	-0.07	0.5 (NS)	-0.02	0.8 (NS)	-0.49	<10e-5	0.18	0.09 (NS)
ΔppFEF ₂₅₋₇₅	0.69	<10e-12	0.12	0.3 (NS)	-0.24	0.02	0.22	0.04	0.63	<10e-10	0.10	0.4 (NS)
ΔST	0.36	<10e-3	-0.31	0.004	0.68	<10e-12	-0.51	<10e-6	0.22	0.04	0.02	0.8 (NS)
ΔATB	0.13	0.4 (NS)	0.56	<10e-7	0.68	<10e-12	0.45	<10e-5	-0.06	0.6 (NS)	-0.01	1 (NS)
Variance (%)	37.5		18.7		17.1		13.3		11.7		1.7	

Corr : Pearson correlation coefficient ; BMI: Body Mass Index ; ppFEV₁: percent predicted Forced Expiratory Volume in 1s ; ppFEF 25-75: percent predicted Forced Expiratory Flow between 25% and 75% of FVC ; ppFVC: percent predicted Forced Vital Capacity ; ATB: days of antibiotics in the last 6 months ; ST: Sweat Test ; NS: not significant

ANNEXE 3

Loading matrix for the 6 outcome parameters and the 6 Principal Components for patients aged 12 years and older.

	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6
Δ BMI Z-score	0.059	-0.78	0.16	0.59	-0.021	-0.096
Δ ppFEV ₁	0.64	0.070	-0.12	-0.061	-0.11	-0.75
Δ ppFVC	0.56	-0.043	-0.067	-0.023	-0.59	0.57
Δ ppFEF ₂₅₋₇₅	0.46	0.11	-0.24	0.24	0.75	0.31
Δ ST	0.24	-0.29	0.67	-0.58	0.26	0.072
Δ ATB	0.084	0.53	0.67	0.51	-0.074	-0.020

BMI: Body Mass Index ; ppFEV₁: percent predicted Forced Expiratory Volume in 1s ; ppFEF₂₅₋₇₅: percent predicted Forced Expiratory Flow between 25% and 75% of FVC ; ppFVC: percent predicted Forced Vital Capacity ; ATB: days of antibiotics in the last 6 months ; ST: Sweat Test.

ANNEXE 4

Mean (min/max) parameter values according to patient clustering in patients aged 12 years and older, and comparison of these clusters according to the studied outcome parameters.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Padj cluster 1/2	Padj cluster 1/3	Padj cluster 1/4	Padj cluster 2/3	Padj cluster 2/4	Padj cluster 3/4
N	10	28	29	21	NA	NA	NA	NA	NA	NA
Age	14.8 (12.2 - 17.3)	14.9 (12.1 - 17.9)	14.4 (12.0 - 18.0)	14.7 (12.0 - 18.5)	NA	NA	NA	NA	NA	NA
BMI M0	-1.0 (-1.9 - 0.04)	-0.6 (-2.3 - 1.1)	-0.2 (-2.2 - 2.5)	-0.8 (-2.7 - 0.44)	0.2 (NS)	0.08 (NS)	0.4 (NS)	0.4 (NS)	0.6 (NS)	0.2 (NS)
ppFEV ₁ M0	62.1 (27.0 - 99.0)	83.9 (25.8 - 127.0)	83.7 (31.4 - 124.0)	91.9 (59.0 - 121.0)	0.08 (NS)	0.08 (NS)	0.02	0.8 (NS)	0.3 (NS)	0.3 (NS)
ppFVC M0	72.5 (31.7 - 108.0)	92.6 (27.7 - 120.0)	93.5 (50.2 - 128.0)	95.8 (61.4 - 142.0)	0.03	0.04	0.03	1 (NS)	0.9 (NS)	0.9 (NS)
ppFEF ₂₅₋₇₅ M0	55.9 (11.0 - 115.6)	68.7 (27.0 - 125.0)	65.5 (14.0 - 124.0)	81.2 (41.0 - 150.6)	0.3 (NS)	0.3 (NS)	0.07 (NS)	1 (NS)	0.3 (NS)	0.3 (NS)
ST M0	90.5 (56.0 - 106.0)	98.7 (73.0 - 126.0)	106.2 (87.0 - 130.0)	98.9 (78.0 - 118.0)	0.2 (NS)	0.002	0.1 (NS)	0.01	0.8 (NS)	0.03
ATB M0	25.5 (0.0 - 60.0)	22.9 (0.0 - 77.0)	19.1 (0.0 - 111.0)	23.8 (0.0 - 105.0)	0.7 (NS)	0.3 (NS)	0.9 (NS)	0.7 (NS)	0.8 (NS)	0.3 (NS)
ΔBMI Z-score	0.3 (-0.3 - 1.5)	0.3 (-0.3 - 0.8)	-0.2 (-0.9 - 0.4)	0.3 (-0.4 - 0.8)	0.7 (NS)	0.02	1 (NS)	<10e-5	0.8 (NS)	<10e-3
ΔppFEV ₁	21.3 (7.4 - 36.1)	5.5 (-2.0 - 17.8)	1.4 (-11.0 - 20.0)	-10.0 (-22.0 - 1.4)	<10e-3	<10e-3	<10e-4	0.04	<10e-6	<10e-4
ΔppFVC	18.8 (1.0 - 36.3)	4.1 (-4.0 - 17.3)	1.4 (-7.1 - 10.2)	-8.1 (-36.0 - 5.2)	0.004	<10e-3	<10e-4	0.2 (NS)	<10e-4	<10e-3
ΔppFEF ₂₅₋₇₅	20.5 (-27.0 - 51.0)	8.4 (-12.0 - 42.0)	3.3 (-33.0 - 45.0)	-17.6 (-39.6 - -4.0)	0.2 (NS)	0.07 (NS)	<10e-3	0.3 (NS)	<10e-5	<10e-4
ΔST	-1.6 (-33.0 - 35.0)	-11.6 (-38.0 - 14.0)	-37.2 (-78.0 - -10.0)	-20.0 (-41.0 - 20.0)	0.3 (NS)	<10e-3	0.02	<10e-4	0.2 (NS)	0.007
ΔATB	10.9 (-20.0 - 44.0)	-16.1 (-56.0 - 0.0)	5.2 (-97.0 - 90.0)	-9.1 (-38.0 - 60.0)	0.004	0.6 (NS)	0.05 (NS)	<10e-3	0.9 (NS)	0.01

BMI: Body Mass Index ; ppFEV1: percent predicted Forced Expiratory Volume in 1s ; ppFEF 25-75: percent predicted Forced Expiratory Flow between 25% and 75% of FVC ; ppFVC: percent predicted Forced Vital Capacity ; ATB: days of antibiotics in the last 6 months ; ST: Sweat Test. Padj: Wilcoxon signed rank test p value adjusted by Benjamini-Hochberg correction ; NA: not applicable ; NS: not significant.

ANNEXE 5

Correlation coefficients and corresponding p values between the 6 outcome parameters and the 6 Principal Components for 6-12 year-old patients. Percentage of explained variance for each PC.

	PC 1		PC 2		PC 3		PC 4		PC 5		PC 6	
	Corr	p	Corr	p	Corr	p	Corr	p	Corr	p	Corr	p
Δ BMI Z-score	0.67	<10e-3	-0.09	0.7 (NS)	0.43	0.04	-0.59	0.002	0.09	0.7 (NS)	-0.01	1 (NS)
Δ ppFEV ₁	0.93	<10e-10	-0.10	0.6 (NS)	-0.20	0.3 (NS)	0.15	0.5 (NS)	-0.09	0.7 (NS)	-0.22	0.3 (NS)
Δ ppFVC	0.81	<10e-5	0.18	0.4 (NS)	-0.46	0.02	-0.08	0.7 (NS)	-0.24	0.3 (NS)	0.16	0.4 (NS)
Δ ppFEF ₂₅₋₇₅	0.78	<10e-5	-0.36	0.09 (NS)	0.23	0.3 (NS)	0.37	0.08 (NS)	0.25	0.2 (NS)	0.11	0.6 (NS)
Δ ST	0.21	0.3 (NS)	0.91	<10e-8	-0.25	0.2 (NS)	-0.01	1 (NS)	0.27	0.2 (NS)	-0.02	0.9 (NS)
Δ ATB	0.19	0.4 (NS)	0.54	0.007	0.77	<10e-4	0.20	0.3 (NS)	-0.21	0.3 (NS)	0.01	1 (NS)
Variance (%)	44.5		21.5		19.1		9.2		4.2		1.5	

Corr : Pearson correlation coefficient ; BMI: Body Mass Index ; ppFEV₁: percent predicted Forced Expiratory Volume in 1s ; ppFEF 25-75: percent predicted Forced Expiratory Flow between 25% and 75% of FVC ; ppFVC: percent predicted Forced Vital Capacity ; ATB: days of antibiotics in the last 6 months ; ST: Sweat Test ; NS: not significant.

ANNEXE 6

Loading matrix for the 6 outcome parameters and the 6 Principal Components for 6-12 year-old patients.

	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6
Δ BMI Z-score	0.41	-0.083	0.40	-0.79	0.19	-0.039
Δ ppFEV ₁	0.57	-0.090	-0.19	0.21	-0.17	-0.75
Δ ppFVC	0.50	0.16	-0.43	-0.11	-0.48	0.55
Δ ppFEF ₂₅₋₇₅	0.48	-0.31	0.22	0.49	0.50	0.37
Δ ST	0.13	0.80	-0.23	-0.0072	0.54	-0.065
Δ ATB	0.12	0.47	0.72	0.28	-0.41	0.018

BMI: Body Mass Index ; ppFEV₁: percent predicted Forced Expiratory Volume in 1s ; ppFEF₂₅₋₇₅: percent predicted Forced Expiratory Flow between 25% and 75% of FVC ; ppFVC: percent predicted Forced Vital Capacity ; ATB: days of antibiotics in the last 6 months ; ST: Sweat Test.

ANNEXE 7

Patients' characteristics at the moment of sweat collection

	Ag e (ye ars)	Sex	Genotype		Genot ype classif ication	Sweat test (mmol/ L)	Pancreati c status	FEV l pp	Bronchial colonization
Patients									
1	13	M	F508del (a)	F508del (a)	2/2	81	PI	104	<i>S.aureus</i>
2	16	M	444del A	444delA	1/1	107	PI	ND	<i>S.aureus</i> , <i>Alcaligenes</i>
3	13	M	F508del (a)	L558S	2/2	105	PI	33	<i>P.aeruginosa</i> , <i>Alcaligenes spp</i>
4	12	F	F508del (a)	F508del(a)	2/2	110	PI	116	<i>S.aureus</i>
5	13	F	G542X(c)	G542X(c)	1/1	118	PI	90	<i>S.aureus</i> , <i>Alcaligenes spp</i>
6	3	M	G542X(c)	G542X(c)	1/1	114	PI	ND	<i>P.aeruginosa</i>
7	17	M	F508del (a)	F508del(a)	2/2	120	PI	129	<i>S.aureus</i>
8	19	F	F508del (a)	F508del(a)	2/2	69	PI	96	<i>S.aureus</i>
9	0.1	M	F508del (a)	W1282X(c)	2/1	95	PI	ND	None detected
10	3	M	Y275X(c)	S466X(c)	1/1	126	PI	ND	<i>S.aureus</i>
11	0.1	F	F508del (a)	del exons 2-3(b)	2/1	102	PI	ND	None detected
12	13	F	F508del (a)	H199Y(a)	2/2	101	PI	109	<i>S.aureus</i>
13	4	M	F508del (a)	F191V(a)	2/2	22	PS	100	None detected
14	3	F	1717-1 G>A(d)	1717-1G >A(d)	1/1	109	PI	ND	None detected
15	5	F	F508del (a)	F508del(a)	2/2	110	PI	ND	None detected
16	13	M	F508del (a)	F508del(a)	2/2	105	PI	93	<i>S.aureus</i>
17	10	M	F508del (a)	F508del(a)	2/2	104	PI	129	<i>S.Maltophilia</i>
18	1.2	M	I507del (a)	Del Exon 17B (b)	2/1	111	PI	ND	None detected
19	0.1	M	F508del (a)	F508del(a)	2/2	96	PI	ND	None detected
20	15	M	2789+5 G>A	3120+1G >A	5/1	103	PS	59	<i>S.aureus</i>
21	17	M	F508del (a)	2143delT	2/1	106	PI	41	<i>P.aeruginosa</i> , <i>Alcaligenes spp</i>
22	0.1	M	F508del (a)	F508del(a)	2/2	98	PI	ND	<i>S.aureus</i>
23	8	F	F508del (a)	F508del(a)	2/2	114	PI	116	None detected

24	14	F	F508del(a)	F508del(a)	2/2	92	PI	108	<i>S.aureus</i>
25	13	F	F508del(a)	R347H(b)	2/2	49	PS	100	None detected
Non-CF control subjects									
1	25	M				25			
2	0.8	M				11			
3	14	M				<10			
4	24	F				10			
5	0.4	F				<10			
6	8	F				17			
7	6	F				17			
8	27	F				12			
9	25	M				19			
10	18	M				12			
11	0.8	M				11			
12	3	M				12			

ANNEXE 8

List of differentially expressed proteins in CF and WT sweat.

Protein	P	Q (Benjamini-Hochberg)	Log Fold change (CF - WT)	Expression in CF sweat
TGM3	1,58E-09	1,57E-06	1,48E+00	Overexpressed
PLBD1	2,19E-08	1,09E-05	3,20E+00	Overexpressed
TGM5	6,44E-08	1,97E-05	2,10E+00	Overexpressed
TOLLIP	1,72E-06	2,84E-04	2,59E+00	Overexpressed
KLK7	2,00E-05	1,32E-03	2,81E+00	Overexpressed
KLK13	1,98E-05	1,32E-03	1,49E+00	Overexpressed
CHI3L1	3,01E-05	1,57E-03	2,63E+00	Overexpressed
ANXA2;ANXA2P2	3,89E-05	1,83E-03	1,92E+00	Overexpressed
CTSB	4,28E-05	1,93E-03	1,06E+00	Overexpressed
CYCS	5,02E-05	2,07E-03	2,22E+00	Overexpressed
CNP	6,03E-05	2,39E-03	1,87E+00	Overexpressed
VCL	9,67E-05	3,09E-03	1,04E+00	Overexpressed
CTBS	1,00E-04	3,10E-03	7,88E-01	Overexpressed
CTSV	1,70E-04	4,00E-03	1,95E+00	Overexpressed
GGH	2,06E-04	4,63E-03	6,91E-01	Overexpressed
SERPINB6	2,18E-04	4,80E-03	8,56E-01	Overexpressed
WDR5	2,60E-04	5,48E-03	2,57E+00	Overexpressed
CHAD	2,66E-04	5,49E-03	2,13E+00	Overexpressed
SNRPN;SNRPB	2,75E-04	5,57E-03	2,28E+00	Overexpressed
CTSA	3,58E-04	6,95E-03	9,90E-01	Overexpressed
LYZ	3,55E-04	6,95E-03	2,70E+00	Overexpressed
SLUR2	4,20E-04	7,84E-03	2,12E+00	Overexpressed
PRCP	4,15E-04	7,84E-03	9,56E-01	Overexpressed

QPCT	4,94E-04	9,07E-03	1,37E+00	Overexpressed
ACP5	5,27E-04	9,33E-03	1,03E+00	Overexpressed
FAM129B	7,94E-08	1,97E-05	-2,08E+00	Underexpressed
RHOA;RHOC	2,04E-07	4,05E-05	-2,41E+00	Underexpressed
VPS4B	9,61E-06	1,19E-03	-2,14E+00	Underexpressed
MPO	1,08E-05	1,19E-03	-3,43E+00	Underexpressed
TFG	9,78E-06	1,19E-03	-1,84E+00	Underexpressed
ELANE	1,51E-05	1,28E-03	-2,48E+00	Underexpressed
GRIPAP1	1,48E-05	1,28E-03	-4,45E+00	Underexpressed
CCT7	1,54E-05	1,28E-03	-2,05E+00	Underexpressed
SLC6A14	1,71E-05	1,30E-03	-2,06E+00	Underexpressed
SLC3A2	2,26E-05	1,40E-03	-8,81E-01	Underexpressed
RPL11	2,67E-05	1,56E-03	-1,92E+00	Underexpressed
P4HB	2,85E-05	1,57E-03	-8,50E-01	Underexpressed
G6PD	3,29E-05	1,63E-03	-3,14E+00	Underexpressed
TMED9	4,50E-05	1,94E-03	-1,87E+00	Underexpressed
HSP90AB1	6,37E-05	2,43E-03	-2,41E+00	Underexpressed
MTHFD1	7,24E-05	2,66E-03	-1,39E+00	Underexpressed
CALR	8,13E-05	2,78E-03	-9,88E-01	Underexpressed
CCT8	7,98E-05	2,78E-03	-2,12E+00	Underexpressed
B3GNT2	8,49E-05	2,80E-03	-1,88E+00	Underexpressed
HSP90B1	1,04E-04	3,14E-03	-2,16E+00	Underexpressed
HSP90AA1	1,16E-04	3,27E-03	-1,24E+00	Underexpressed
SLC38A10	1,14E-04	3,27E-03	-1,84E+00	Underexpressed
RPS18	1,21E-04	3,33E-03	-1,68E+00	Underexpressed
SLC25A5;SLC25A6	1,27E-04	3,40E-03	-1,88E+00	Underexpressed
MFGE8	1,35E-04	3,51E-03	-1,58E+00	Underexpressed
RAB11A;RAB11B	1,49E-04	3,79E-03	-1,07E+00	Underexpressed

PSMD6	1,54E-04	3,82E-03	-8,88E-01	Underexpressed
DSG3	1,70E-04	4,00E-03	-1,44E+00	Underexpressed
SLC9A3R1	1,92E-04	4,42E-03	-2,15E+00	Underexpressed
MB	2,40E-04	5,18E-03	-1,53E+00	Underexpressed
FAM129A	5,17E-04	9,31E-03	-1,83E+00	Underexpressed
NCALD;HPCA;HP CAL1	5,54E-04	9,64E-03	-1,30E+00	Underexpressed

ANNEXE 9

Differentially expressed proteins in delF508 cells treated or not with the tritherapy (TRI)

Student's T-test p-value DMSO_TRI	Student's T-test q-value DMSO_TRI	Expression in TRI	Student's T-test Test statistic DMSO_TRI	Protein_names	Gene_names
2,10E-06	1,21E-02	Over	-1,41E+01	Thiosulfate sulfurtransferase	TST
8,65E-06	2,00E-02	Over	-1,15E+01	Propionyl-CoA carboxylase beta chain, mitochondrial	PCCB
1,05E-05	2,00E-02	Over	-1,11E+01	Glutathione S-transferase kappa 1	GSTK1
1,79E-05	2,57E-02	Over	-1,03E+01	Electron transfer flavoprotein subunit alpha, mitochondrial	ETFA
2,35E-05	2,69E-02	Over	-9,86E+00	Carnitine O-palmitoyltransferase 1, liver isoform	CPT1A
3,52E-05	3,37E-02	Under	9,27E+00	Epidermal growth factor receptor kinase substrate 8-like protein 1	EPS8L1
5,92E-05	3,40E-02	Over	-8,56E+00	Fumarate hydratase, mitochondrial	FH
5,81E-05	3,40E-02	Over	-8,58E+00	Enoyl-CoA hydratase, mitochondrial	ECHS1
5,03E-05	3,40E-02	Over	-8,77E+00	Succinyl-CoA ligase [ADP/GDP-forming] subunit alpha, mitochondrial	SUCLG1
4,59E-05	3,40E-02	Over	-8,90E+00	Succinyl-CoA ligase [GDP-forming] subunit beta, mitochondrial	SUCLG2
7,44E-05	3,76E-02	Under	8,26E+00	Wolframin	WFS1
8,04E-05	3,76E-02	Over	-8,16E+00	ATP synthase subunit beta, mitochondrial	ATP5B
8,51E-05	3,76E-02	Over	-8,09E+00	Lanosterol 14-alpha demethylase	CYP51A1
1,20E-04	4,39E-02	Over	-7,67E+00	Calcium-binding mitochondrial carrier protein Aralar1	SLC25A12
2,51E-04	4,39E-02	Over	-6,81E+00	3-ketoacyl-CoA thiolase, peroxisomal	ACAA1
2,45E-04	4,39E-02	Over	-6,84E+00	Pyruvate dehydrogenase E1 component subunit beta, mitochondrial	PDHB
2,12E-04	4,39E-02	Over	-7,00E+00	Ferrochelatase, mitochondrial	FECH
1,18E-04	4,39E-02	Over	-7,68E+00	Isovaleryl-CoA dehydrogenase, mitochondrial	IVD
1,46E-04	4,39E-02	Over	-7,42E+00	Stress-70 protein, mitochondrial	HSPA9
2,52E-04	4,39E-02	Over	-6,81E+00	Enoyl-CoA delta isomerase 1, mitochondrial	ECI1
1,94E-04	4,39E-02	Over	-7,10E+00	Leucine-rich PPR motif-containing protein, mitochondrial	LRPPRC
1,64E-04	4,39E-02	Over	-7,29E+00	Short/branched chain specific acyl-CoA dehydrogenase, mitochondrial	ACADSB

2,16E-04	4,39E-02	Over	-6,98E+00	Elongation factor Tu, mitochondrial	TUFM
1,58E-04	4,39E-02	Over	-7,33E+00	Isocitrate dehydrogenase [NAD] subunit alpha, mitochondrial	IDH3A
2,06E-04	4,39E-02	Over	-7,03E+00	Trifunctional enzyme subunit beta, mitochondrial;3-ketoacyl-CoA thiolase	HADHB
1,72E-04	4,39E-02	Over	-7,24E+00	Heat shock protein 75 kDa, mitochondrial	TRAP1
1,68E-04	4,39E-02	Over	-7,27E+00	NAD(P) transhydrogenase, mitochondrial	NNT
1,88E-04	4,39E-02	Over	-7,14E+00	A-kinase anchor protein 1, mitochondrial	AKAP1
2,03E-04	4,39E-02	Over	-7,05E+00	Protein TBRG4	TBRG4
2,49E-04	4,39E-02	Under	6,82E+00	Homer protein homolog 3	HOMER3
2,53E-04	4,39E-02	Over	-6,80E+00	Mycophenolic acid acyl-glucuronide esterase, mitochondrial	ABHD10
2,24E-04	4,39E-02	Over	-6,94E+00	Succinyl-CoA ligase [ADP-forming] subunit beta, mitochondrial	SUCLA2
2,12E-04	4,39E-02	Over	-7,00E+00	AFG3-like protein 2	AFG3L2
2,66E-04	4,49E-02	Over	-6,75E+00	39S ribosomal protein L48, mitochondrial	MRPL48
2,87E-04	4,71E-02	Over	-6,66E+00	ATP synthase subunit alpha, mitochondrial	ATP5A1
3,08E-04	4,91E-02	Over	-6,59E+00	Acyl-CoA dehydrogenase family member 9, mitochondrial	ACAD9
3,18E-04	4,94E-02	Over	-6,55E+00	Adenylate kinase 2, mitochondrial;Adenylate kinase 2, mitochondrial, N-terminally processed	AK2

ANNEXE 10

List of Overdispersed and Shifted Modules according to rRoma

Shifted	Overdispersed
IL6_JAK_STAT3_SIGNALING	G2M_CHECKPOINT
ADIPOGENESIS	PI3K_AKT_MTOR_SIGNALING
PROTEIN_SECRETION	
UNFOLDED_PROTEIN_RESPONSE	
MTORC1_SIGNALING	
MYC_TARGETS_V1	
XENOBIOTIC_METABOLISM	
FATTY_ACID_METABOLISM	
OXIDATIVE_PHOSPHORYLATION	
HALLMARK_GLYCOLYSIS	
REACTIVE_OXIGEN_SPECIES_PATHWAY	
UV_RESPONSE_DN	
HALLMARK_PEROXISOME	

ORIGINAL ARTICLE



Profiling the response to lumacaftor-ivacaftor in children with cystic between fibrosis and new insight from a French-Italian real-life cohort

Matthieu Cornet PhD¹ | Geneviève Robin PhD² | Fabiana Ciciriello PhD³ |
 Tiphaine Bihoue MD, PhD⁴ | Christophe Marguet MD⁵ | Valérie Roy⁶ |
 Muriel Lebourgeois MD⁵ | Frédérique Chedevergne MD⁵ |
 Anne Sophie Bonnel MD⁵ | Mairead Kelly PhD¹ | Philippe Reix MD, PhD⁷ |
 Vincenzina Lucidi MD³ | Véronique Stoven PhD⁶ |
 Isabelle Sermet-Gaudelus MD, PhD¹

¹Institut Necker-Enfants Malades, Paris, France

²Laboratoire de Mathématiques et Modélisation d'Evry, Evry, France

³Ospedale Pediatrico Bambino Gesù, Roma, Italy

⁴Centre Hospitalier Universitaire de Nantes, Nantes, France

⁵Hôpital Necker Enfants Malades, Paris, France

⁶Mines Paris – PSL, Paris, France

⁷Hospices Civils de Lyon, Lyon, France

Correspondence

Isabelle Sermet-Gaudelus, Hôpital Necker Enfants Malades, 149 rue de Sévres, 75015 Paris, France.
 Email: isabelle.sermet@aphp.fr

Funding information

Fondation pour la Recherche Médicale

Abstract

Introduction: Clinical trials for CFTR modulators consider mean changes of clinical status at the cohort level, and thus fail to assess the heterogeneity of the response. We aimed to study the different response profiles to lumacaftor-ivacaftor according to age in children with cystic fibrosis (CF).

Methods: A mathematical framework, including principal component analysis, data clustering, and data completion, was applied to a multicenter cohort of 112 children aged 6–18 years, treated with lumacaftor-ivacaftor. Studied parameters at baseline and 6 months included body mass index (BMI), number of days of antibiotics (ATB), Sweat test (ST), forced expiratory volume in 1 s expressed in percentage predicted (ppFEV₁), forced vital capacity (ppFVC), and forced expiratory flow at 25%–75% of FVC (ppFEF_{25–75}).

Results: Change in ppFEV₁ was the most significant parameter in characterizing response heterogeneity among the 12–18-year-old patients. Patients with minimal changes in ppFEV₁ were further separated by change in BMI and ATB course. In the 6–12-year-old children both BMI and ppFEV₁ evolution were the most relevant. ST change was not associated with a clinical response.

Conclusions: Change in ppFEV₁, BMI, and ATB course are the most relevant outcomes to discriminate clinical response profiles in children treated with lumacaftor-ivacaftor. Prepubertal and pubertal children display different response profiles.

Vincenzina Lucidi, Véronique Stoven, and Isabelle Sermet-Gaudelus are co-last authors.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. *Pediatric Pulmonology* published by Wiley Periodicals LLC.

KEYWORDS

children, lumacaftor-ivacaftor, principal component analysis

1 | INTRODUCTION

CFTR modulators are expected to rapidly improve baseline condition of people with cystic fibrosis (CF). Clinical trials and real-world studies show improvement of pulmonary function tests (PFTs).^{1–3} Other outcomes are also considered, including sweat chloride concentration, pulmonary exacerbations, and nutritional status. Importantly, these parameters cannot be considered independently as they all reflect the consequences of CFTR dysfunction on different target tissues. We aimed to determine which combination of outcomes, collected on a routine clinical basis, might be the most relevant to assess the global heterogeneity of the response to CFTR modulators in children treated with lumacaftor-ivacaftor, and to characterize different profiles of response based on the evolution of these clinical outcomes in a 6-months span.

We focused on a cohort of children older than 6 years treated with the combo therapy lumacaftor-ivacaftor and followed-up in different pediatric centers in France and in Italy. We considered the parameters usually measured in patients treated with CFTR modulators, including PFT, body mass index (BMI), sweat test (ST), and antibiotic courses, and studied their evolution during the first 6 months of treatment. Independent parameters characterizing the response variability were identified with principal component analysis (PCA). PCA provides a mathematical framework to integrate various parameters, identifies principal components (PCs) that explain most of the variability in the data, which variables contribute the most to each component, and whether they vary together. It was also used as a dimension reduction method. We took advantage of this functionality in our study, as k-means clustering was performed on the minimum number of PCs explaining 70% of the variance of the sample.

2 | MATERIALS AND METHODS

2.1 | Cohorts and data description

This study is a retrospective analysis of data collected during routine follow-up of patients. Data from children above 6 years homozygous for p.Phe508del and treated with lumacaftor-ivacaftor were collected from four centers: Necker Enfants Malades, Paris, France; Ospedale pediatrico Bambino Gesù, Rome, Italy; Hospices Civils de Lyon, Lyon, France; and Centre Hospitalier Universitaire de Nantes, Nantes, France. Patients were enrolled before starting lumacaftor-ivacaftor therapy and were evaluated at 6 months according to the standard of care. The following outcomes were measured at treatment initiation (M0) and at six months (M6), according to the standard of care guidelines: forced expiratory volume in 1 s (FEV₁), forced vital capacity (FVC), and forced expiratory flow at 25%–75%

of FVC (FEF_{25–75}), all expressed as percentage predicted (pp) values for normal values⁴; chloride concentration by ST; days of antibiotics in the previous 6 months as a treatment of pulmonary exacerbation, including both IV treatment and oral therapy (ATB); BMI, expressed as BMI Z-score.⁵ Clinical data were part of a French and Italian real-life study (IRB Comité d'Ethique de la Recherche APHP.5, nber 0011928; IRB Ospedale Pediatrico Bambino Gesù, nber 1534).

We studied their absolute change between M0 and M6: Δ ppFEV₁, Δ ppFVC, Δ ppFEF_{25–75}, Δ BMI, Δ ATB, Δ ST where “ Δ ” stands for “absolute variation from baseline to 6 months.”

All patients initiating lumacaftor-ivacaftor combination were enrolled in the study. Patients were included between 2017 and 2020.

2.2 | Principal component analysis (PCA) algorithm and imputation of missing data

We used PCA to describe our data set.⁶ If patients are described by n initial clinical parameters, PCA determines n new variables that are linear combinations of the initial parameters, called principal components (PC) 1 to n (PC1 to PC n). The PCs define n new orthogonal (i.e., independent or non-redundant) variables along which patients are best dispersed, such that the percentage of variance captured along the corresponding directions decreases from PC1 to PC n . Initial parameters with strong coefficients in the same PC are correlated. Conversely, initial parameters that never appear together with strong coefficients in any component are essentially independent. Using PCA, it is also possible to view the PCs as linear combinations of the initial parameters, showing which parameters contribute the most to each PC. This information is available in loading matrixes: the higher the loading of an initial parameter in a given PC, the more it contributes to this PC, and therefore, to the variance along this PC direction. Loadings are similar to correlations, which can also be calculated and represented graphically to visualize relationships between initial parameters and PCs.⁷

In our study, six PCs were determined, named PC1 to PC6. As PCs are ordered by the percentage of overall variance explained, PC1 describes the major part of the variance of the data, followed by PC2, and so forth. Thus, parameters included in the first PCs are the most relevant to discriminate patient response. In particular, those associated with PC1 are the most important. On the contrary, the last PCs carry little to no information and can be regarded as uninformative. In this study, we thus only considered the PCs explaining most of the global variance (at least two-thirds of the variance when taken together). Their correlation with the six outcome parameters was studied and represented graphically in two-dimensional (2D) space to ease interpretation. Each vector (represented by an arrow) represents an initial parameter, that is, one

of the monitored outcomes. The more a vector is aligned with a PC, the longer the arrow, the more it is correlated to the PC, and thus contributes to the variance captured by this PC.

To remove potential bias of PCA towards parameters with larger variations, data for each outcome parameter were normalized after aggregation, before running the algorithm. Missing values were imputed before running the PCA algorithm, using the missForest R package.⁸

2.3 | Patient clustering

Clustering was performed by Kmeans algorithm. This method allows to separate the samples in clusters of equal variances and to provide an informative and interpretable description of the data. The optimal number of clusters was determined using the C-index criterion.⁹ The clusters were compared based on the six outcome variables considered in the study to better characterize the associated profiles of response and understand what differentiate them from each other. To take advantage of the dimension reduction feature of PCA, patient clustering was performed based on the minimum number of PCs necessary to account for 70% of the variance of the sample.

2.4 | Statistical evaluation

Analyses were conducted in R.¹⁰ Data are presented as mean (IQR). A *p* value of <0.05 was considered statistically significant. As normality assumption could not be verified for all parameters, comparisons between values from patients in the 6–12-year-old cohort and values from patients in the 12–18-year-old cohort, as well as comparisons between values from patients in different subgroups of the same age group, were performed by Wilcoxon signed-rank test. Values from patients in the same age group at baseline and after 6 months of treatment were compared by paired Wilcoxon signed-rank test. Correlation between variables was assessed by Spearman correlation test. Benjamini–Hochberg corrections were applied to account for multiple testing.

3 | RESULTS

3.1 | Patient characteristics

Characteristics of the 112 F508del homozygous children are summarized in Supporting Information: Table 1. Patients had a mild lung disease, as assessed by a mean ppFEV₁ of 87% at M0. Similarly, most of the patients had a normal nutritional status with a mean BMI Z-score at M0 of –0.6. Six- to 12 year-old patients (*n* = 24) had a significantly better respiratory status than 12–18-year-old patients (*n* = 88), as shown by a significantly higher ppFEV₁ (*p* < 10e–3) and ppFEF_{25–75} (*p* = 0.01). Therefore, the two cohorts were analyzed separately.

Overall, in both cohorts, there was a significant improvement in BMI and ST at 6 months lumacaftor-ivacaftor, in contrast to ppFEV₁ and ppFEF_{25–75} which were not significantly changed.

Forty-two values were missing in the 12–18-year-old cohort (4.0% of the data) and 9 in the 6–12-year-old cohort (3.1%). Most of them involved the ST. Removing all patients with at least one missing value would have led to discard 43 patients in the PCA analysis. We thus performed missing values imputation, as described in Supporting Information. No significant change in data interpretation was observed upon potential data imputation errors after performing bootstrap-based multiple imputations using repeated random patient removal, which provides evidence for the robustness of the data imputation method¹¹ (Supporting Information: Table 1 and Figure 1A).

3.2 | PCA of clinical parameters and patient clustering in the 12–18 years children cohort

PCA analysis determined six PCs, PC1 to PC6. PCA was robust to data reduction, as shown by PCA calculation after repeated random data removal (Supporting Information: Figure 1B). The amount of variance explained by each component is shown in Supporting Information: Table 2. We focused on the first three PCs in the following analysis (PC1, PC2, and PC3), as they accounted for more than 70% of the whole variance of the sample (respectively 38%, 19%, and 17%; Supporting Information: Table 2).

Evolution of the three PFT parameters (Δ ppFEV₁, Δ ppFVC, Δ ppFEF_{25–75}) were the strongest contributors to PC1, as shown by their correlation with this PC (Figure 1 and Supporting Information: Table 2) and their value in the loading matrix (Supporting Information: Table 3). Δ ppFVC and Δ ppFEF_{25–75} carried redundant information with Δ ppFEV₁, as assessed by their similar loadings in PC1 (Supporting Information: Table 3). Δ ST was also correlated to PC1, although to a lesser degree, as shown by less alignment of its vector to PC1 and a smaller loading (Figure 1 and Supporting Information: - Tables 2 and 3). Δ ATB and Δ BMI were strongly correlated to PC2 (Figure 1 and Supporting Information: Table 2). As expected, these two parameters pointed in opposite directions; for example, when BMI increased, the number of ATB courses in the six previous months decreased. Δ ST and Δ ATB were the most significant contributors to PC3 (Figure 1 and Supporting Information: Table 2).

Clustering was then performed to characterize profiles of response to treatment. Patient clustering was performed based on the first three principal components. An optimal number of four clusters was determined by the C-index criterion. Two-dimensional (2D) visualization of the patients within clusters in the PC space is shown in Figures 2 and 3 and Supporting Information: Figure 2. Partition of Cluster 1 and Cluster 4 was based on respiratory outcomes. Cluster 1 was characterized by an improvement in all three pulmonary function parameters (Figure 3 and Supporting Information: Table 4). Δ ppFEV₁ change in Cluster 1 was significantly higher than in any other cluster, as indicated by an absolute increase of 21%

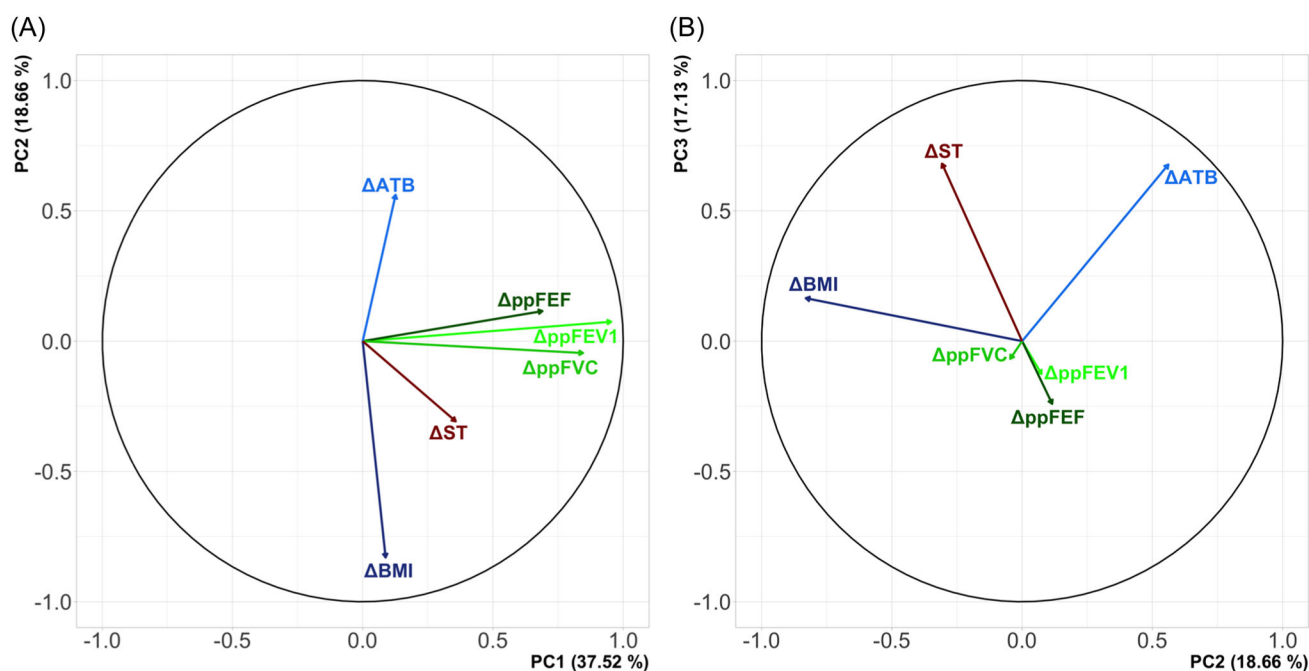


FIGURE 1 Circle of correlation between clinical parameters and the first three PCA components for patients aged 12 years and older. (A) Graphical representation of PC1 and PC2. (B) Graphical representation of PC2 and PC3. Each parameter is represented by a vector whose abscissa and ordinate provide its correlation with PCA components. % of explained overall variance of the data are displayed alongside the PC they correspond to. ATB, days of antibiotics in the last 6 months; BMI, body mass index; ppFEF, percent predicted forced expiratory flow between 25% and 75% of FVC; PC, principal components; PCA, principal component analysis; ppFEV₁, percent predicted forced expiratory volume in 1 s; ppFVC, percent predicted forced vital capacity; ST, sweat test.

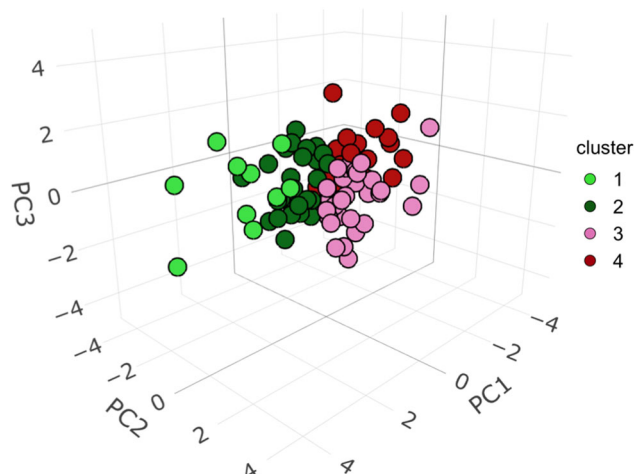


FIGURE 2 Projection of the patients on the three-dimensional space with coordinates of the first three principal components (PC1, PC2, and PC3). Each point corresponds to one patient and each color corresponds to a cluster. Coordinates for each patient were obtained by a linear combination of the outcome parameters contributing to the PCs, according to PCA methodology. PC, principal components; PCA, principal component analysis.

(9) points, while it was significantly lower in Cluster 4 than in any other cluster (absolute decrease of 10% (6) points) (Supporting Information: Table 4). A similar pattern was observed for Δ ppFVC or

Δ ppFEF₂₅₋₇₅, although to a lesser degree. In particular, Δ ppFEF₂₅₋₇₅ was not significantly higher in Cluster 1 compared to Clusters 2 or 3. A combination of Δ ppFEV₁ with Δ ppFVC and Δ ppFEF₂₅₋₇₅ within PC1 (Figure 3) did not improve the separation between clusters, further supporting the fact that Δ ppFEV₁, Δ ppFVC, and Δ ppFEF₂₅₋₇₅ carry redundant information. Interestingly, patients from Cluster 1 had a significantly lower ppFVC at treatment initiation compared to any other cluster, and a significantly lower ppFEV₁, as compared to Cluster 4. Patients in Cluster 1 also displayed the smallest decrease in ST (-1.6 (20)). Those in Clusters 2 and 3 displayed an intermediate change in respiratory parameters and were further differentiated by their respective change in ATB course, BMI Z-score, and ST. Patients in Cluster 2 displayed the greatest decrease in ATB. Patients in Cluster 3 had a significantly higher ST at M0 than any other cluster, displayed the greatest decrease in ST (-37 (18)) and, in contrast to all other clusters, experienced a decrease in BMI Z-score.

3.3 | PCA of clinical parameters and patient clustering in the 6–12 years children cohort

In children below 12 years of age, the two first components of the PCA summarized 66% of the total variance, PC1 alone accounting for 44% (Supporting Information: Table 5). We therefore focused on the first two PCs. Given the small number of patients, we did not perform clustering analysis.

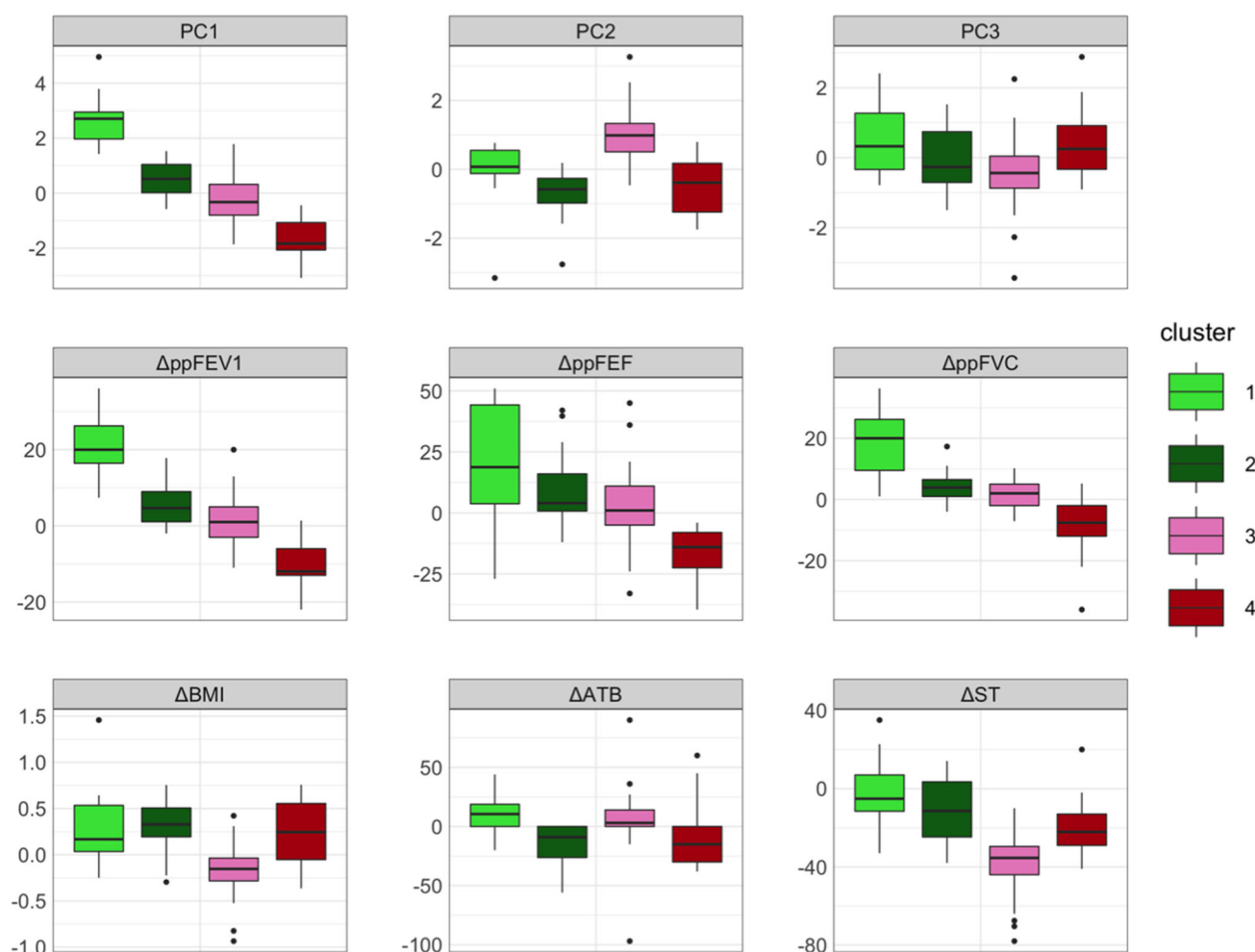


FIGURE 3 Distribution of parameters in the four clusters for 12–18-year-old patients. Values for the first three PCs were obtained by a linear combination of the outcome parameters contributing to the PCs, according to PCA methodology. ATB, days of antibiotics in the last 6 months; BMI, body mass index; PC, principal components; PCA, principal component analysis; ppFEF, percent predicted forced expiratory flow between 25% and 75% of FVC; ppFEV₁, percent predicted forced expiratory volume in 1 s; ppFVC, percent predicted forced vital capacity; ST, sweat test.

Similarly to the 12–18-year-old cohort, PC1 was correlated to the three respiratory function parameters (Figure 4 and Supporting Information: Table 5). Δ ppFEF and Δ ppFVC remained highly correlated with Δ ppFEV₁, as assessed by their similar loadings (Supporting Information: Table 6). Conversely, unlike the 12–18-year-old cohort, Δ BMI was also a strong contributor to PC1 ($r = 0.67$, $p = 0.0003$) (Supporting Information: Table 5). Δ ST and Δ ATB were the two main contributors to PC2 (Figure 4 and Supporting Information: Tables 5 and 6).

4 | DISCUSSION

This study aimed at characterizing the heterogeneity of the response to lumacaftor-ivacaftor in children enrolled in a “real-life” study and at deciphering different profiles of response to treatment. We used a mathematical framework based on PCA to

identify nonredundant variables and random forests to impute missing values.

This also enabled to show the relationships between the evolution of outcome parameters in a 6-month span, highlighting specific response patterns. Nutritional outcome, as assessed by BMI, was an important response parameter in the 6–12-year-old patients, in contrast to the 12–18-year-old cohort.

Given the efficiency of CFTR modulators, an increasing number of children will benefit from these therapies as it is anticipated that if started early in life, they might at least slow down the evolution of the disease.¹² However, the fact that children present a very mild disease or are even asymptomatic makes evaluation in real-life studies very challenging.¹³ This is why outcomes such as Lung Clearance Index are currently being tested to characterize the response.¹⁴ However, such investigations cannot be performed in all centers. We investigated whether outcomes collected routinely in all European centers

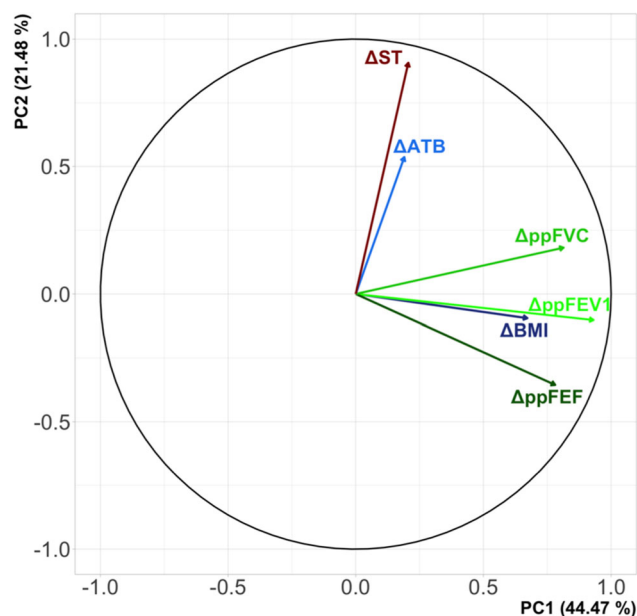


FIGURE 4 Circle of correlation between clinical parameters and the first two PCA components, for patients younger than 12 years old. Each parameter is represented by a vector whose abscissa and ordinate provide its correlation with PCA components. % of explained overall variance of the data are displayed alongside the PC they correspond to. ATB, days of antibiotics in the last 6 months; BMI, body mass index; PC, principal components; PCA, principal component analysis; ppFEF, percent predicted forced expiratory flow between 25% and 75% of FVC; ppFEV₁, percent predicted forced expiratory volume in 1 s; ppFVC, percent predicted forced vital capacity; ST, sweat test.

and in a standardized way might differentiate specific response patterns. In contrast to previous studies where the effect of a treatment was evaluated in terms of a mean response of the whole cohort, we focused on the heterogeneity of the response in real-life follow-up. We used PCA to reduce the number of variables to a smaller number of principal components without losing too much information and for later clustering. We assessed the redundancy between the outcome parameters to define a core data set of the response of the pediatric patients. Therefore, our analysis differs from those performed usually by understanding which parameters vary similarly in response to a treatment. Not only did we assess which factors are implicated in cohort heterogeneity, but we also aimed at finding a combination of variables that explain most of the variability in a given data set. Moreover, we defined different profiles of response thanks to clustering analysis. To our knowledge, a single other study focused on the heterogeneity of the response to treatment in patients with bronchiectasis, and also concluded that the response was highly heterogeneous.¹⁵ These observations reinforce the need to understand how multiple factors participate to the response.

Importantly, the methodology we propose in this study can be applied to other studies. In particular, following the advent of

the triple combination therapy to treat patients with CF,¹⁶ it would be interesting to apply this methodology to a cohort of patient treated with elexacaftor-tezacaftor-ivacaftor.

The main limitation of our study is the small number of patients. We thus challenged the statistical soundness of our results based on a random resampling approach, assessing how much PCA results would have changed upon removal of part of our data. In addition, there were missing data in this real-life study. Although only a few results were missing (around 4%), ignoring patients with partial information would have led to discard around 38% of the patients, which is critical in already small cohorts. We implemented a methodology of data imputation and assessed whether the results of the PCA could change with respect to potential data imputation errors. We found that the PCA results were stable regarding both resampling and potential imputation errors. Therefore, our methodology is statistically robust despite the small size of the cohort and our results are reliable.

Another limitation is the absence of follow-up beyond 6 months. Although this time range is similar to the one used in clinical studies^{17,18} and allows to unveil significant heterogeneity in response to treatment, the possible long-term effects of the treatment and long-term heterogeneity cannot be assessed by the current study.

In the 12–18-year-old cohort, clustering of the patients unveiled two response levels. First, even if it was not significant at the whole sample level, the evolution of ppFEV₁ allowed to characterize “Good Respiratory Responders” with significant improvement (Cluster 1), “Bad Respiratory Responders” with significant deterioration (Cluster 4) and clusters with minimal change, within the intrinsic variability of this respiratory test.¹⁹ This provides evidence that ppFEV₁ evolution is of crucial significance to discriminate response of pediatric patients aged 12–18 years treated with lumacaftor-ivacaftor. Moreover, neither ΔppFVC nor ΔppFEF_{25–75} improved this cluster characterization, suggesting that they were redundant with ΔppFEV₁ and did not provide additional information. Interestingly, patients in Cluster 1 (those who improved the most) had the worse ppFEV₁ at baseline and a significantly lower ppFVC. This suggests that in a pediatric population, patients with lower PFT, and in particular ppFEV₁, have a higher potential for improvement. This is supported by similar data from adolescents, which showed an improvement of ppFEV₁ in those with low baseline ppFEV₁,^{20,21} while increase in ppFEV₁ was not detected in patients with normal ppFEV₁.^{1,22} Overall, ppFEV₁ evolution in this cohort underlines the importance of considering subgroups of patients. Indeed, considering the mean evolution of ppFEV₁ at the sample level would have led to the conclusion that there is no significant improvement. On the contrary, the subgroups analysis highlights a significant improvement in specific subpopulations of the cohort and underlines the heterogeneity of the response to the treatment. The patients benefiting the most were those with a significantly lower ppFVC at treatment initiation compared to any other cluster and a significantly lower ppFEV₁ as compared to Cluster 4.

To further characterize the response of the patients with no significant change in ppFEV₁ (Clusters 2 and 3), the second level

of response testing (i.e., PC2) included change in ST, BMI Z-score, and ATB courses. Cluster 2 was characterized by an increase in BMI and a decrease in ATB consumption (i.e., less pulmonary exacerbations). This highlights a specific profile of patients experiencing concomitant improved nutrition and infection in contrast to minimal change in ppFEV₁. The emergence of ATB courses as an independent factor of the response in pediatric cohorts was also evidenced in other cohorts with normal ppFEV₁.^{23,24} This may reflect improved pulmonary inflammation and innate immunity, based on improved antimicrobial activity and immune cell function.^{25,26} This highlights the potential benefit of those therapies even in patients with normal lung function. This aspect is crucial in the pediatric population who displays a “silent” disease in a perspective of preventative care. Similar to our cohort, a significant improvement in BMI has been reported in the 12–18-year-old children treated with lumacaftor-ivacaftor.^{1,20} This may reflect not only an increased caloric intake but also an improved pancreatic status, and possibly a systemic effect of increased anabolism due to improved inflammation.²⁷ Unexpectedly, we also identified a group of patients who worsened their BMI (Cluster 3). We have no clear explanation for this, but the fact that these patients displayed neither improvement in ppFEV₁ nor in ATB course shows that they were not clinically improved at all by the modulator therapy. Interestingly, these patients also displayed the most important decrease in ST, while patients with the best respiratory response displayed the smallest decrease in ST. This demonstrates that, even though sweat chloride concentration improves at the level of the overall cohort, its evolution is not indicative of a clinical benefit at the individual level in the pediatric population. This is in line with previous studies in adults showing that sweat chloride concentration decreases in response to treatment with CFTR modulators, although there was no correlation with ppFEV₁ evolution.^{27–29} Patients in this cluster had a significantly higher ST value than those in any other cluster, thus they might have had a larger margin of improvement for this parameter.

The 6–12-year-old cohort was analyzed separately because they had a significantly better respiratory function at baseline as compared to the 12–18-year-old adolescents ($p = 0.002$). In this younger cohort, ppFEV₁ did not significantly improve, as already reported in studies conducted with lumacaftor-ivacaftor and tezacaftor-ivacaftor in a similar age group (6–11-year-olds).^{17,18,22} Most importantly, the PCA profile was different from the 12–18-year-old cohort, as PC1 was not only correlated to change in ppFEV₁ but also in BMI. This shows that in this younger sample with normal respiratory function, BMI captures a part of the response and is an important parameter, independently of the respiratory function. This is indirectly confirmed by the fact that in most of the studies in this age group, BMI is consistently improved in contrast to ppFEV₁.^{17,30} BMI evolution is usually considered as a secondary endpoint. Here, we show that BMI evolution is an important endpoint in young patients to discriminate responders from nonresponders.

5 | CONCLUSION

We propose a global mathematical framework suited to real-life studies that could be translated to currently ongoing studies.

In the present case, it highlighted that changes in ppFEV₁, BMI, and ATB course numbers explain most of the heterogeneity in children's response to lumacaftor-ivacaftor. In particular, BMI is the cornerstone of the heterogeneity of the response in the 6–12-year-old children with a normal respiratory function. Improvements in lung function, BMI, exacerbation frequency, and ST are dissociated in children with CF. Studying heterogeneity of the response will help reassess how we measure treatment response to CFTR modulators in this population, who will be the target of the future trials with CFTR modulators.

AUTHOR CONTRIBUTIONS

Matthieu Cornet: Conceptualization; methodology; software; visualization; validation; writing – original draft; formal analysis. **Geneviève Robin:** Methodology; conceptualization, writing – original draft. **Fabiana Ciciriello:** Data curation; writing – review & editing; investigation. **Tiphaine Bihouee:** Data curation; investigation; writing – review & editing. **Christophe Marguet:** Data curation; investigation; writing – review & editing. **Valérie Roy:** Conceptualization; methodology. **Muriel Lebourgeois:** Investigation; writing – review & editing; data curation. **Frédérique Chedevergne:** Data curation; investigation. **Anne Sophie Bonnel:** Data curation; investigation; writing – review & editing. **Mairead Kelly:** Writing – review & editing. **Philippe Reix:** Data curation; writing – review & editing; investigation. **Vincenzina Lucidi:** Data curation; supervision; writing – review & editing; investigation. **Véronique Stoven:** Conceptualization; methodology; supervision; writing – original draft. **Isabelle Sermet-Gaudelus:** Writing – original draft; resources; data curation; project administration; supervision; conceptualization; investigation; funding acquisition; writing – review & editing; methodology.

ACKNOWLEDGMENTS

We thank Fabien Perrin for data collection. We thank Giuliana Lancellotti for data collection at the Ospedale Pediatrico Bambino Gesù, Rome, Italy. We thank ABCF association for the support for the study. Matthieu Cornet is supported by the Fondation pour la Recherche Médicale.

CONFLICT OF INTEREST

ISG has participated to scientific Board of Vertex Therapeutics and received Academic Funding from Vertex Therapeutics. The other authors do not declare any conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Matthieu Cornet  <http://orcid.org/0000-0001-5466-9373>

Philippe Reix  <http://orcid.org/0000-0002-9192-8335>

Isabelle Sermet-Gaudelus  <https://orcid.org/0000-0001-5537-9482>

REFERENCES

- Sagel SD, Khan U, Heltshe SL, et al. Clinical effectiveness of lumacaftor/ivacaftor in patients with cystic fibrosis homozygous for F508del-CFTR. A clinical trial. *Ann ATS*. 2021;18(1):75-83.
- Burgel P-R, Munck A, Durieu I, et al. Real-life safety and effectiveness of lumacaftor-ivacaftor in patients with cystic fibrosis. *Am J Respir Crit Care Med*. 2020;201(2):188-197.
- Corey M, Edwards L, Levison H, Knowles M. Longitudinal analysis of pulmonary function decline in patients with cystic fibrosis. *J Pediatr*. 1997;131(6):809-814.
- Quanjer PH, Stanojevic S, Cole TJ, et al. Multi-ethnic reference values for spirometry for the 3-95-yr age range: the global lung function 2012 equations. *Eur Respir J*. 2012;40(6):1324-1343.
- World Health Organization. *WHO Child Growth Standards: Length/Height-for-Age, Weight-for-Age, Weight-for-Length, Weight-for-Height, and Body Mass Index-for Age: Methods and Development*. 1st ed; 2006.
- Jolliffe IT. *Principal Component Analysis*. Springer; 1986.
- Lê S, Josse J, Husson F. FactoMineR: an R package for multivariate analysis. *J Stat Softw*. 2008;25(1). <http://www.jstatsoft.org/v25/i01/>
- Stekhoven DJ, Bühlmann P. MissForest--non-parametric missing value imputation for mixed-type data. *Bioinformatics*. 2012;28(1):112-118.
- Hubert LJ, Levin JR. A general statistical framework for assessing categorical clustering in free recall. *Psychol Bull*. 1976;83(6):1072-1080.
- Ihaka R, Gentleman R. R: a language for data analysis and graphics. *J Comput Graph Stat*. 1996;5(3):299.
- Efron B. Missing data, imputation, and the bootstrap. *J Am Stat Assoc*. 1994;89(426):463-475.
- Accurso FJ. Treatment of cystic fibrosis in infants. *Lancet Respir Med*. 2018;6(7):483-484.
- Stanojevic S, Ratjen F. Physiologic endpoints for clinical studies for cystic fibrosis. *J Cyst Fibros*. 2016;15(4):416-423.
- Ratjen F, Klingel M, Black P, et al. Changes in Lung Clearance Index in preschool-aged patients with cystic fibrosis treated with Ivacaftor (GOAL): a clinical trial. *Am J Respir Crit Care Med*. 2018;198(4):526-528.
- Sibila O, Laserna E, Shoemark A, et al. Heterogeneity of treatment response in bronchiectasis clinical trials. *Eur Respir J*. 2021;59:2100777.
- Zemanick ET, Taylor-Cousar JL, Davies J, et al. A phase 3 open-label study of elxacaftor/tezacaftor/ivacaftor in children 6 through 11 years of age with cystic fibrosis and at least one F508del allele. *Am J Respir Crit Care Med*. 2021;203(12):1522-1532.
- Ratjen F, Hug C, Marigowda G, et al. Efficacy and safety of lumacaftor and ivacaftor in patients aged 6-11 years with cystic fibrosis homozygous for F508del-CFTR: a randomised, placebo-controlled phase 3 trial. *Lancet Respir Med*. 2017;5(7):557-567.
- Walker S, Flume P, McNamara J, et al. A phase 3 study of tezacaftor in combination with ivacaftor in children aged 6 through 11 years with cystic fibrosis. *J Cyst Fibros*. 2019;18(5):708-713.
- Cooper PJ, Robertson CF, Hudson IL, Phelan PD. Variability of pulmonary function tests in cystic fibrosis. *Pediatr Pulmonol*. 1990;8(1):16-22.
- Wainwright CE, Elborn JS, Ramsey BW, et al. Lumacaftor-ivacaftor in patients with cystic fibrosis homozygous for Phe508del CFTR. *N Engl J Med*. 2015;373(3):220-231.
- Taylor-Cousar JL, Munck A, McKone EF, et al. Tezacaftor-ivacaftor in patients with cystic fibrosis homozygous for Phe508del. *N Engl J Med*. 2017;377(21):2013-2023.
- Milla CE, Ratjen F, Marigowda G, Liu F, Waltz D, Rosenfeld M. Lumacaftor/ivacaftor in patients aged 6-11 years with cystic fibrosis and homozygous for F508del-CFTR. *Am J Respir Crit Care Med*. 2017;195(7):912-920.
- Stick SM. Clinical trials in infants with cystic fibrosis detected following newborn screening. *Paediatr Respir Rev*. 2008;9(3):176-180.
- Britto MT, Kotagal UR, Hornung RW, Atherton HD, Tsevat J, Wilmott RW. Impact of recent pulmonary exacerbations on quality of life in patients with cystic fibrosis. *Chest*. 2002;121(1):64-72.
- Khan MA, Ali ZS, Sweezey N, Grasemann H, Palaniyar N. Progression of cystic fibrosis lung disease from childhood to adulthood: neutrophils, neutrophil extracellular trap (NET) formation, and NET degradation. *Genes*. 2019;10(3):183.
- Geitani R, Moubareck CA, Xu Z, Karam Sarkis D, Touqui L. Expression and roles of antimicrobial peptides in innate defense of airway mucosa: potential implication in cystic fibrosis. *Front Immunol*. 2020;11:1198.
- Barry PJ, Jones AM, Webb AK, Horsley AR. Sweat chloride is not a useful marker of clinical response to Ivacaftor. *Thorax*. 2014;69(6):586-587.
- Durmowicz AG, Witzmann KA, Rosebraugh CJ, Chowdhury BA. Change in sweat chloride as a clinical end point in cystic fibrosis clinical trials. *Chest*. 2013;143(1):14-18.
- Masson A, Schneider-Futschik EK, Baatallah N, et al. Predictive factors for lumacaftor/ivacaftor clinical response. *J Cyst Fibros*. 2019;18(3):368-374.
- Davies JC, Wainwright CE, Canny GJ, et al. Efficacy and safety of ivacaftor in patients aged 6 to 11 years with cystic fibrosis with a G551D mutation. *Am J Respir Crit Care Med*. 2013;187(11):1219-1225.

SUPPORTING INFORMATION

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

How to cite this article: Cornet M, Robin G, Ciciriello F, et al. Profiling the response to lumacaftor-ivacaftor in children with cystic between fibrosis and new insight from a French-Italian real-life cohort. *Pediatric Pulmonology*. 2022;57:2992-2999. doi:10.1002/ppul.26123

ORIGINAL ARTICLE

Proteomic profiling of sweat in patients with cystic fibrosis provides new insights into epidermal homeostasis

Matthieu Cornet^{1,2,3,4}  | Thao Nguyen-Khoa^{1,5} | Mairead Kelly-Aubert^{1,6} | Vincent Jung⁷ | Frédérique Chedevergne^{8,9} | Muriel Le Bourgeois^{8,9} | Laura Aoust^{8,9} | Kévin Roger⁷ | Chiara Ida Guerrero⁷ | Isabelle Sermet-Gaudelus^{1,6,8,9}

¹Institut Necker Enfants Malades, INSERM U1151, Paris, France

²Center for Computational Biology, Mines ParisTech, PSL Research University, Paris, France

³Institut Curie, Paris, France

⁴INSERM U900, Paris, France

⁵Laboratory of Biochemistry, Hôpital Universitaire Necker Enfants Malades AP-HP Centre, Paris, France

⁶Université Paris Cité, Paris, France

⁷Proteomics Platform Necker, Université Paris Cité, Structure Fédérative de Recherche Necker (SFR Necker, INSERM US24/CNRS UAR3633), Paris, France

⁸Centre de Référence Maladies Rares, Mucoviscidose et Maladies Apparentées, Hôpital Necker Enfants Malades AP-HP Centre Paris Cité, Paris, France

⁹European Respiratory Network, ERN Lung, Frankfurt, Germany

Correspondence

Isabelle Sermet-Gaudelus, Institut Necker Enfants Malades, INSERM U1151, Paris, France; Université Paris Cité, France; Centre de Référence Maladies Rares, Mucoviscidose et Maladies Apparentées, Hôpital Necker Enfants Malades, AP-HP Centre Paris Cité, Paris, France; European Respiratory Network, ERN Lung, Frankfurt Germany.
Email: isabelle.sermet@aphp.fr

Funding information

Mucoviscidose ABCF2

Abstract

Background: A high proportion of patients with Cystic Fibrosis (CF) also present the rare skin disease aquagenic palmoplantar keratoderma. A possible link between this condition and absence of a functional CF Transmembrane conductance Regulator protein in the sweat acinus and collecting duct remains unknown.

Methods: In-depth characterization of sweat proteome profiles was performed in 25 CF patients compared to 12 healthy controls. A 20 µL sweat sample was collected after pilocarpine iontophoresis and liquid chromatography tandem mass spectrometry (LC-MS/MS) proteomic analysis was performed.

Results: Sweat proteome profile of CF patients was significantly different from that of healthy subjects with 57 differentially expressed proteins. Cystic Fibrosis sweat proteome was characterized by an increase in 25 proteins including proteases (Kallikrein 7 and 13, Phospholipase B domain containing 1, Cathepsin A L2 and B, Lysosomal Pro-X carboxypeptidase); proinflammatory proteins (Annexin A2, Chitinase-3-like protein 1); cytochrome c and transglutaminases. Thirty-two proteins were downregulated in CF sweat including proteases (Elastase 2), antioxidative protein FAM129 B; membrane-bound transporter SLC6A14 and regulator protein Sodium-hydrogen antiporter 3 regulator 1.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. Skin Health and Disease published by John Wiley & Sons Ltd on behalf of British Association of Dermatologists.

Conclusion: This study is the first to report in-depth characterization of endogenous peptides in CF sweat and could help understand the complex physiology of the sweat gland. The proteome profile highlights the unbalanced proteolytic and proinflammatory activity of sweat in CF. These results also suggest a defect in pathways involved in skin barrier integrity in CF patients. Sweat proteome profile could prove to be a useful tool in the context of personalized medicine in CF.

1 | INTRODUCTION

One of the hallmarks of Cystic Fibrosis (CF) is high Chloride (Cl^-) sweat concentration due to the absence of the CF Transmembrane conductance Regulator (CFTR) protein in the sweat acinus and collecting duct.¹ Patients with CF (pwCF) can display aquagenic palmo-plantar keratoderma (AKP), characterized by exaggerated and early wrinkling of the skin after brief immersion in water associated with tingling and pain sensation.² This condition is probably underestimated as AKP is observed in 41%–84% of CF patients.^{3,4} The physiopathology of AKP and its relationship to CFTR dysregulation is not yet fully elucidated. This condition has been recently reported to be improved by CFTR modulators in CF patients which relates this skin phenotype to CFTR defects.⁵ To gain further insight into the pathophysiology of this rare skin disorder, we performed in-depth analysis of the sweat proteome of pwCF compared to non-CF individuals.

2 | PATIENTS AND METHODS

2.1 | Patients

We enrolled 25 pwCF, and 12 non-CF individuals (WT). Patients' genotype was classified according to genetic mechanism of the variant (class 1: nonsense, deletion or severe splicing; class 2: misfolding; class 3: activation defect; class 4: conductance defect; class 5: splicing with residual function).¹ All participants signed an informed consent form for sweat collection and analysis (no. Eudract 2016-A00309-42; Comité de Protection des Personnes IDF2).

Clinical status of the patient at the study visit was assessed by pancreatic status, microbiology of sputum and Respiratory Volume in 1 s (FEV1), expressed as percentage predicted (ppFEV1). Sweat induction was performed by pilocarpine iontophoresis, as previously reported.⁶ Sweat was collected using the Wescor Macroduct (ELITech, Puteaux, France) and chloride concentration was assessed by coulometry (chloridometer 926S Sherwood, Servilab, Le Mans, France). Twenty μL of remaining sweat sample was used for proteomic analysis.

What is already known about this topic?

- Although sweat is a reliable, non-invasive, and easy to collect biofluid, sweat proteome profiles as a source of biomarkers has been studied in relatively few diseases.

What does this study add?

- Our study is the first to attempt a comparative analysis of sweat proteome profiles between Cystic Fibrosis (CF) patients and controls and reveals a unique subset of differentially expressed proteins. The results provide new insights to epidermal homeostasis in CF.

What is the translational message?

- Sweat proteome profiles could serve as convenient tools in CF for diagnosis or personalized therapeutic interventions.

2.2 | Sample preparation for liquid chromatography mass spectrometry (LC-MS/MS) analysis

Twenty μL of sweat were processed in the S-TrapTM micro spin column (Protifi, Hutington, USA) for digestion according to manufacturer's instructions. Briefly, sodium dodecyl sulphate (SDS) was added to a final concentration of 5% and samples were reduced with 20 mM tris(2-carboxyethyl)phosphine (TCEP) and alkylated with 50 mM chloroacetamide (CAA) for 15 min at room temperature. Aqueous phosphoric acid was then added to a final concentration of 1.2% followed by the addition of six volumes of S-Trap binding buffer (90% aqueous methanol, 100 mM triethylammonium bicarbonate buffer (TEAB), pH 7.1). Mixtures were then loaded on S-Trap columns. Six washing steps were performed for thorough SDS elimination. Samples were digested with 0.8 μg of trypsin (Promega) at 47°C for 1h30. After elution, peptides were vacuum dried and resuspended in 30 μL of 2% acetonitrile (ACN), 0.1% formic acid (FA) in HPLC-grade water prior to mass

spectrometry (MS) analysis. The concentration of the peptide mixtures was determined using a Nanodrop 2000 from Labtech France.

2.3 | LC-MS/MS analysis

For each sample, 400 ng of peptides were injected on a nanoelute (Bruker Daltonics, Germany) high-performance liquid chromatography system coupled to a timsTOF Pro (Bruker Daltonics, Germany) mass spectrometer. High-performance liquid chromatography separation (Solvent A: 0.1% FA in water, 2% ACN; Solvent B: 0.1% FA in ACN) was carried out at 250 nL/min using a packed emitter column (C18, 25 cm × 75 µm 1.6 µm) (Ion Optics, Australia) with a gradient elution (2%–11% solvent B during 19 min; 11%–16% for 7 min; 16%–25% for 4 min; 25%–80% for 3 min and finally 80% for 7 min to wash the column). Mass-spectrometric data were acquired using the parallel accumulation serial fragmentation (PASEF) acquisition method. The measurements were carried out over the *m/z* range from 100 to 1700 Th. The ion mobilities values ranged from 0.8 to 1.3 V s/cm² (1/k0). The total cycle time was set to 1.16 s and the number of PASEF MS/MS scans was set to 10. For low sample amounts, the total cycle time was set to 1.88 s. QCs, consisting in 10 ng of Hela total cell lysates were run before, after and during the acquisition of the series.

2.4 | Data analysis

The data were analyzed using MaxQuant version 2.0.17.0 and searched with Andromeda search engine against the UniProtKB/Swiss-Prot Homo sapiens database (release 02-2021, 20 396 entries). To search parent mass and fragment ions, we set a mass deviation of 3 and 20 ppm respectively. The minimum peptide length was set to 7 amino acids and strict specificity for trypsin cleavage was required, allowing up to two missed cleavage sites. Carbamidomethylation (Cys) was set as fixed modification, whereas oxidation (Met) and N-term acetylation were set as variable modifications. The false discovery rates at the protein and peptide level were set to 1%. Scores were calculated in MaxQuant as described previously.⁷ The reverse and common contaminants hits were removed from MaxQuant output. Proteins were quantified according to the MaxQuant label-free algorithm using label free quantification intensities, providing a label-free normalization; protein quantification was obtained using at least 2 peptides per protein. Match between runs was allowed.

Data filtering and imputation was performed using the Prostar Software. Proteins were retained in the analysis if they were detected in at least 70% of the patients in at least one group. Statistical analyses were

conducted in R. Data in pwCF and WT conditions were compared by Student *t*-test. Benjamini-Hochberg corrections were applied to account for multiple testing. A *q*-value <0.01 combined with a log2 fold change >0.5 was considered statistically significant.

3 | RESULTS

3.1 | Population

We enrolled 25 pwCF, mean age 9.1(6.3) years, all carrying 2 CF causing variants of the *CFTR* gene (11 homozygous p.Phe508del (F508del thereafter); 7 F508del compound heterozygotes; 7 carrying nonsense, splicing mutations or large deletion variants) (Table 1). Patient 13 carried the rare missense variant p.Phe191Val (F191 V) in *trans* of F508del and had a normal chloride sweat concentration at 22 mmol/L. However, β -adrenergic sweat secretion rate was in the CF range, providing insight for a genotype with minimal CFTR activity.⁸ Patient 20 carried the 2789 + 5G > A splice site mutation, known to be associated with a small amount of normally spliced transcripts.⁹ All patients were at steady state at the study visit. Three patients (patient 13, 20, 25) were pancreatic sufficient. Patients 2, 3, and 21 had chronic sputum colonization with *Pseudomonas.aeruginosa* or *Alcaligenes.spp* and were treated with inhaled antibiotics. The 12 non-CF individuals (WT) had a mean age of 12.5 (10.8) years and a normal sweat test.

3.2 | Proteomic analysis

Protein concentration ranged between 0.25 and 0.70 µg/µL of sweat with no difference between pwCF and WT. Nine hundred ninety-one proteins were retained for the analysis. Data are available via ProteomeXchange with identifier PXD032894.

Fifty-seven proteins were found differentially abundant between pwCF and WT subjects: 25 were upregulated in CF sweat, and 32 were downregulated (Figure 1, Table 2). Hierarchical analysis confirmed that these proteins clearly separated the pwCF from the WT as shown in Figure 2.

Sweat proteome profile of F508del homozygous patients was not significantly different from that of compound F508del heterozygotes or patients with other genotypes, as shown in the dendrogram (Figure 2). Furthermore, patients carrying 2 class 1 variants had a similar profile to those carrying at least one class 2 variant. Neither age, nor sex influenced clustering, including in the infants. We thus performed statistical analysis on the overall pwCF cohort to increase the significance of our results.

Cystic Fibrosis sweat proteome was characterized by an increased level in (i) proteases (Kallikrein 7 and

TABLE 1 Patients characteristics at the moment of sweat collection

Age (years)		Sex	Genotype		Genotype classification	Sweat test (mmol/L)	Pancreatic status	FEV1 pp	Bronchial colonization
Patients									
1	13	M	F508del (a)	F508del (a)	2/2	81	PI	104	<i>S.aureus</i>
2	16	M	444delA	444delA	1/1	107	PI	ND	<i>S.aureus</i> , <i>Alcaligenes</i>
3	13	M	F508del(a)	L558S	2/2	105	PI	33	<i>P.aeruginosa</i> , <i>Alcaligenes</i> spp
4	12	F	F508del(a)	F508del(a)	2/2	110	PI	116	<i>S.aureus</i>
5	13	F	G542X(c)	G542X(c)	1/1	118	PI	90	<i>S.aureus</i> , <i>Alcaligenes</i> spp
6	3	M	G542X(c)	G542X(c)	1/1	114	PI	ND	<i>P.aeruginosa</i>
7	17	M	F508del(a)	F508del(a)	2/2	120	PI	129	<i>S.aureus</i>
8	19	F	F508del(a)	F508del(a)	2/2	69	PI	96	<i>S.aureus</i>
9	0.1	M	F508del(a)	W1282X(c)	2/1	95	PI	ND	None detected
10	3	M	Y275X(c)	S466X(c)	1/1	126	PI	ND	<i>S.aureus</i>
11	0.1	F	F508del(a)	del exons 2-3(b)	2/1	102	PI	ND	None detected
12	13	F	F508del(a)	H199Y(a)	2/2	101	PI	109	<i>S.aureus</i>
13	4	M	F508del(a)	F191 V(a)	2/2	22	PS	100	None detected
14	3	F	1717-1G>A(d)	1717-1G>A(d)	1/1	109	PI	ND	None detected
15	5	F	F508del(a)	F508del(a)	2/2	110	PI	ND	None detected
16	13	M	F508del(a)	F508del(a)	2/2	105	PI	93	<i>S.aureus</i>
17	10	M	F508del(a)	F508del(a)	2/2	104	PI	129	<i>S.Maltophilia</i>
18	1.2	M	I507del(a)	Del Exon 17B (b)	2/1	111	PI	ND	None detected
19	0.1	M	F508del(a)	F508del(a)	2/2	96	PI	ND	None detected
20	15	M	2789 + 5G > A	3120 + 1G > A	5/1	103	PS	59	<i>S.aureus</i>
21	17	M	F508del(a)	2143delT	2/1	106	PI	41	<i>P.aeruginosa</i> , <i>Alcaligenes</i> spp
22	0.1	M	F508del(a)	F508del(a)	2/2	98	PI	ND	<i>S.aureus</i>
23	8	F	F508del(a)	F508del(a)	2/2	114	PI	116	None detected
24	14	F	F508del(a)	F508del(a)	2/2	92	PI	108	<i>S.aureus</i>
25	13	F	F508del(a)	R347H(b)	2/2	49	PS	100	None detected
Non-CF control subjects									
1	25	M	NA	NA	NA	25	NA	NA	NA
2	0.8	M	NA	NA	NA	11	NA	NA	NA
3	14	M	NA	NA	NA	<10	NA	NA	NA
4	24	F	NA	NA	NA	10	NA	NA	NA
5	0.4	F	NA	NA	NA	<10	NA	NA	NA
6	8	F	NA	NA	NA	17	NA	NA	NA
7	6	F	NA	NA	NA	17	NA	NA	NA
8	27	F	NA	NA	NA	12	NA	NA	NA
9	25	M	NA	NA	NA	19	NA	NA	NA
10	18	M	NA	NA	NA	12	NA	NA	NA
11	0.8	M	NA	NA	NA	11	NA	NA	NA
12	3	M	NA	NA	NA	12	NA	NA	NA

Abbreviations: CF, Cystic Fibrosis; NA, not adapted; ND, not determined; PI, pancreatic insufficient; PS, pancreatic sufficient.



FIGURE 1 Volcano plot of differentially expressed proteins between patients with cystic Fibrosis (CF) (patients with CF (pwCF)) and non-CF individuals (WT). Thresholds of 0.5 for log2 Fold Change (FC) in abscissa, and of 0.01 for q values (expressed as $-\log_{10}$) in ordinate, are used for significance and plotted in green. Differentially expressed proteins are plotted in colour (Proteases: green; proinflammatory proteins: orange; transglutaminases: purple; antioxydative proteins: pink; others: blue)

13, Phospholipase B domain containing 1, Cathepsin A L2 and B, Lysosomal Pro-X carboxypeptidase, neuraminidase); (ii) proinflammatory proteins (Annexin A2, Chitinase-3-like protein 1); (iii) cytochrome c; (iv) Transglutaminases 3 and 5; and (v) Toll-interacting protein (TOLLIP). This contrasted with decrease of Elastase 2, an epidermal protease, FAM129 B, an antioxydative protein and Desmoglein-3. Finally, 2 proteins known to interfere with CFTR biology were underexpressed (Sodium-hydrogen antiporter 3 regulator 1 [NHERF1] and Solute Carrier Family 6 Member 14, SLC6A14).

The severity of the disease, based on a ppFEV1 below 50%, and/or chronic colonization with *Pseudomonas.aeruginosa* or *Alcaligenes spp* did not modify the proteome profile.

4 | DISCUSSION

This study is to our knowledge the first report of CF sweat proteome. We were able to yield nearly 1000 proteins, in as low as 20 μ L of sweat thanks to an optimized protocol with high-resolution LC-MS/MS acquisition. To our knowledge, a similar depth of proteome coverage was

only reported in 2 previous studies in healthy adults and using larger volumes of sweat. Yu et al performed the analysis on 10 ml pooled samples collected from absorbing tissue pads at different body sites after exercise while Burat et al used a mean volume of 70 μ L of sweat.^{10,11} We used 20 μ L of leftover sweat sample collected for the sweat test. Usually at least 50 μ L of sweat are collected for sweat test and around 20 μ L are used for sweat Cl^- concentration evaluation. Our results demonstrate that the remainder aliquot of a diagnostic sweat test can be used reliably for sweat proteomics, including in infants as young as 1 month of age.

A large range of proteases and their respective inhibitors, antimicrobial peptides and effectors of skin innate immunity have already been reported in sweat of healthy subjects^{10–14} or in specific diseases such as schizophrenia or tuberculosis.^{15,16} Detection of other proteins such as cytoskeletal proteins or proteins involved in oxidative stress, unfolded protein response, Endoplasmic Reticulum stress, proteasome and cancellous metabolism pathways points to the diversity of sweat biology.^{11,17–20} Although some studies report a correlation between blood versus sweat concentration for specific biomarkers such as glucose,^{16,21} the literature shows mixed results for others²² and relatively

TABLE 2 List of differentially expressed proteins in cystic fibrosis (CF) and WT sweat

Protein	P	Q (Benjamini-Hochberg)	Log fold change (CF - WT)	Expression in CF sweat
TGM3	1.58170503179399e-09	1.56746968650784e-06	1.48299833333333	Overexpressed
PLBD1	2.19449916269758e-08	1.08737433511665e-05	3.19911266666667	Overexpressed
TGM5	6.43668943899884e-08	1.96835309247718e-05	2.09593033333334	Overexpressed
TOLLIP	1.71984520823241e-06	0.000284061100226386	2.593015	Overexpressed
KLK7	1.99882588608743e-05	0.00132055763540843	2.80988066666667	Overexpressed
KLK13	1.97919945229845e-05	0.00132055763540843	1.48745466666667	Overexpressed
CHI3L1	3.01262586508373e-05	0.0015713222752525	2.6340845	Overexpressed
ANXA2; ANXA2P2	3.88823861049168e-05	0.00183487831571298	1.92317566666667	Overexpressed
CTSB	4.28375399575546e-05	0.00192963645899712	1.05712433333333	Overexpressed
CYCS	5.0216366740295e-05	0.00207351747665135	2.224024	Overexpressed
CNP	6.03328049753096e-05	0.00239159238922127	1.86662933333333	Overexpressed
VCL	9.66636492911453e-05	0.00309011859508145	1.04174433333333	Overexpressed
CTBS	0.000100169383365191	0.00310212059109076	0.788042000000001	Overexpressed
CTSV	0.000169602967955179	0.00400182241056147	1.95409066666667	Overexpressed
GGH	0.00020565187231108	0.00463184103318819	0.690882000000002	Overexpressed
SERPINB6	0.000218078527698848	0.00480257379887907	0.856213666666665	Overexpressed
WDR5	0.000259717676076137	0.00547617482960537	2.56654883333333	Overexpressed
CHAD	0.000265784565267186	0.00548734383707879	2.12672166666667	Overexpressed
SNRPN; SNRPB	0.000275203341900658	0.0055658471800725	2.27634633333333	Overexpressed
CTSA	0.000357792614031708	0.0069524015785377	0.989661000000002	Overexpressed
LYZ	0.000355199902106944	0.0069524015785377	2.700541	Overexpressed
SLUR2	0.000419516866382318	0.00784417386009203	2.11938466666667	Overexpressed
PRCP	0.000415002068705056	0.00784417386009203	0.955500666666666	Overexpressed
QPCT	0.000493990855610998	0.00906564699834257	1.365012	Overexpressed
ACP5	0.000527198325269628	0.00932952750611073	1.02944833333333	Overexpressed
FAM129B	7.94491661948408e-08	1.96835309247718e-05	-2.0791504	Underexpressed
RHOA; RHOC	2.04192488811702e-07	4.04709512824794e-05	-2.41087553333333	Underexpressed
VPS4B	9.60588941329161e-06	0.00119159328624749	-2.144275	Underexpressed
MPO	1.08217351929641e-05	0.00119159328624749	-3.42752986666667	Underexpressed
TFG	9.78268215062519e-06	0.00119159328624749	-1.835101	Underexpressed
ELANE	1.51021002126614e-05	0.00127553183077237	-2.48330053333333	Underexpressed
GRIPAP1	1.47731114847263e-05	0.00127553183077237	-4.4502426	Underexpressed
CCT7	1.54453904836211e-05	0.00127553183077237	-2.04755293333333	Underexpressed
SLC6A14	1.70723579325206e-05	0.00130143897777907	-2.0631152	Underexpressed
SLC3A2	2.25999565002928e-05	0.00139978480573689	-0.881463	Underexpressed
RPL11	2.66865246950978e-05	0.00155566741016717	-1.91789953333333	Underexpressed
P4HB	2.84976534262211e-05	0.00156895414141028	-0.849515333333335	Underexpressed
G6PD	3.29256052077666e-05	0.00163146373804484	-3.14162133333333	Underexpressed
TMED9	4.5015504560557e-05	0.00193958108780487	-1.8657086	Underexpressed
HSP90AB1	6.36883655989668e-05	0.00242750655032985	-2.41141233333333	Underexpressed

TABLE 2 (Continued)

Protein	P	Q (Benjamini-Hochberg)	Log fold change (CF - WT)	Expression in CF sweat
MTHFD1	7.2379130902605e-05	0.0026565821749808	-1.390594	Underexpressed
CALR	8.12542575545831e-05	0.00277665411160662	-0.987771333333333	Underexpressed
CCT8	7.98292581291651e-05	0.00277665411160662	-2.11998133333333	Underexpressed
B3GNT2	8.49078785657414e-05	0.00280479025528832	-1.879466	Underexpressed
HSP90B1	0.000104409061495325	0.00313543575581414	-2.160834	Underexpressed
HSP90AA1	0.00011554029444987	0.00327144090856632	-1.23706566666667	Underexpressed
SLC38A10	0.000113663277186564	0.00327144090856632	-1.84155993333333	Underexpressed
RPS18	0.000120845307526837	0.00332660277108598	-1.67556433333333	Underexpressed
SLC25A5; SLC25A6	0.000126964753433738	0.00340059650413066	-1.8838592	Underexpressed
MFGE8	0.000134504290978555	0.00350773032525652	-1.57676266666667	Underexpressed
RAB11A; RAB11B	0.000149194674859679	0.00379107494322929	-1.07444733333333	Underexpressed
PSMD6	0.000154226015316499	0.00382094952946626	-0.887891	Underexpressed
DSG3	0.000169543652295849	0.00400182241056147	-1.4409494	Underexpressed
SLC9A3R1	0.000191808840825127	0.00442052468041165	-2.15198906666667	Underexpressed
MB	0.000240247633526712	0.00517576967010808	-1.53099033333333	Underexpressed
FAM129 A	0.000516733099906237	0.00931059094558329	-1.82837733333333	Underexpressed
NCALD; HPCA; HPCAL1	0.000554437456083902	0.0096394301575289	-1.301822	Underexpressed

Abbreviations: CALR, Calreticulin; CHAD, Chondroadherin; CNP, 2',3'-cyclic nucleotide 3' phosphodiesterase; CTBS, Chitinase; CTSA, Cathepsin A; CTSB, Cathepsin B; CTSV, Cathepsin V; CYCS, Cytochrome C; ELANE, Elastase, Neutrophil Expressed; GGH, Gamma-Glutamyl Hydrolase; HPCA, Hippocalcin; LYZ, Lysozyme C; MB, Myoglobin; MPO, Myeloperoxidase; NCALD, Neurocalcin Delta; PRCP, Prolylcarboxypeptidase; QPCT, Glutaminyl-Peptide Cyclotransferase; RHOA, Ras Homolog Family Member A; RHOC, Ras Homolog Family Member C; SNRPN, small nuclear ribonucleoprotein polypeptide; TFG, Trafficking From ER To Golgi Regulator; VCL, Vinculin.

few proteins have yet been fully assessed for blood versus sweat concentration correlation.

As in previous studies, we did not report any strong influence of sex or age, even in 1 month old babies. Nor did we observe any influence of disease severity based on genetics or lung disease pattern, although the very limited number of patients with severe disease precludes a definitive conclusion.

Our results add additional insights in sweat physiology by showing an unbalanced proteolytic and proinflammatory activity of this biofluid, in pwCF. A number of proteases were upregulated, including Kallikreins 7 and 13. These epidermal-specific serine proteases activate a proteolytic cascade, generating skin inflammation and ultimately skin physical barrier impairment, as already reported in the Netherton Syndrome or Atopic Dermatitis.²³ Annexin A2 forms a complex with S100A10 involved in a large variety of pathophysiological processes, including activation of metalloprotease induced inflammation.²⁴ Other regulators of innate immunity were also significantly upregulated, such as Chitinase-3-like protein (CHI3L1) and TOLLIP. CHI3L1 reflects induction of T-helper cell type 2, IL-13-induced inflammation, and macrophage activation.²⁵ Interestingly, this

protein is also elevated in sputum and serum of CF patients with lung disease.²⁶

Toll-interacting protein is an intermediate in interleukin (IL)-1 signalling involved in the Toll-like receptor mediated inflammation and triggers the nuclear factor (NF)- κ B and mitogen-activated protein kinase signalling.²⁷ All these proinflammatory signalling cascades may be associated with an increased oxidative status, as assessed by the increase in Cytochrome C, a mitochondrial protein involved in the electron transport system and oxidative phosphorylation. This may be potentiated by the decrease of the antioxidative protein FAM129B.²⁸

Our results also suggest a defect in the integrity of the skin barrier in pwCF. Indeed, downregulation of elastase 2 is anticipated to decrease the cleavage of filaggrin, a protein which plays a key role in the integrity of the epidermal skin barrier through its end products.^{29,30} Absence of filaggrin in null carriers predisposes to ichthyosis vulgaris, and/or atopic dermatitis.^{31,32} Downregulation of Desmoglein 3, a component of desmosomes, should reinforce the skin barrier defect by decreasing cell to cell junction.³³ Finally, transglutaminases overexpression may perturb the formation

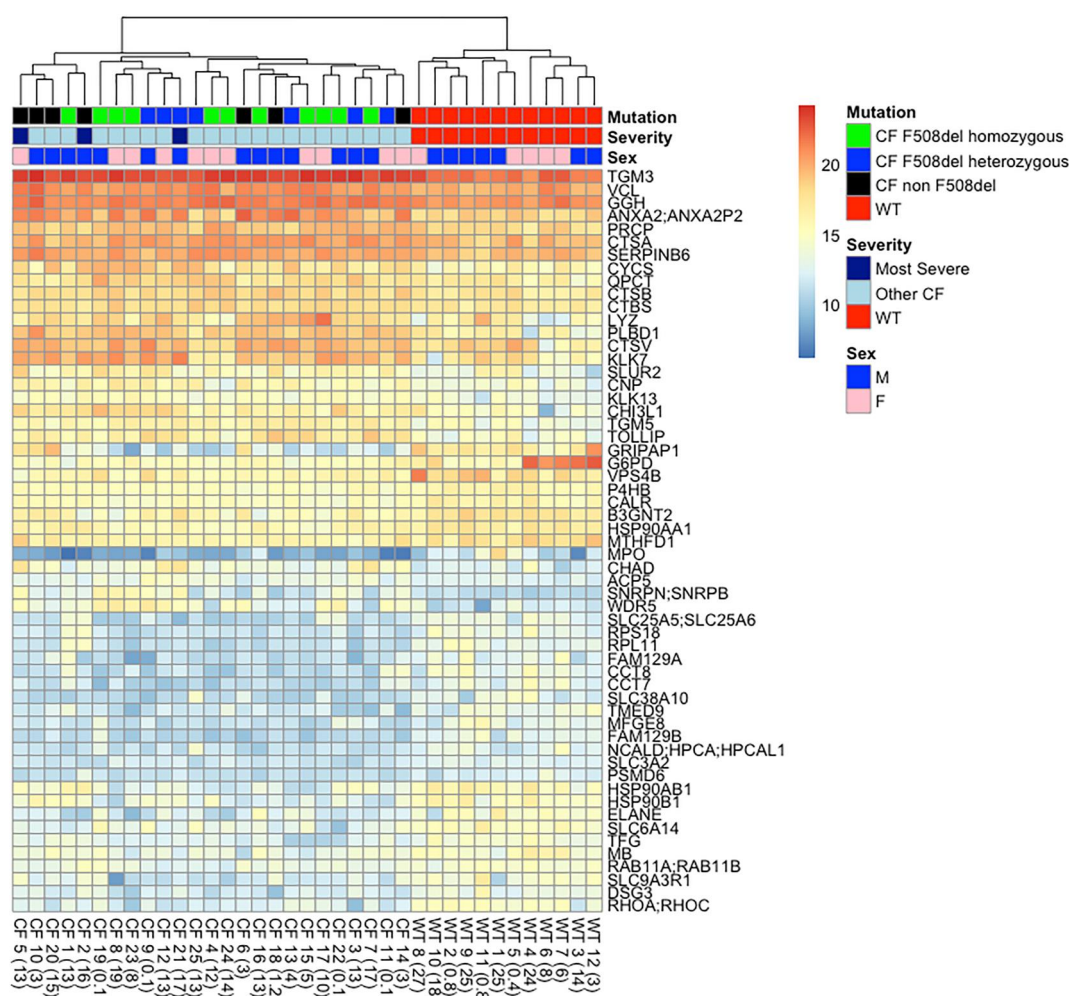


FIGURE 2 Heat map of differentially expressed proteins between patients with cystic fibrosis (CF) (patients with CF (pwCF)) and non-CF individuals (WT). Dendrogram is shown according to genetic status (mutation), severity (most severe: ppFEV1 < 50%, chronic sputum colonization with *P.aeruginosa* or *Alcaligenes* spp), and sex. Patients' age (in years) is presented after their assigned study number

of the outer layer of the epidermis by altering cross-linking of keratins and structural proteins.³⁴

All these defects should result in altered skin barrier function and favour increased water content when the skin is immersed in water, one of the main features observed in AKP.³⁵ Indeed, although the pathogenesis of AKP is not fully understood, it is assumed to be related to hydropic changes of the horny layer and the excessive electrolyte content of CF sweat which moistens the epidermis obviously plays an additional role.³⁴ The observation that CFTR function restoration by ivacaftor, a highly efficient CFTR modulator correcting Cl^- channel activity, improves AKP symptoms provides clinical evidence that this condition is related to CFTR dysfunction.⁵ Interestingly, in our data set, we also show a defective expression of 2 proteins known to modulate the expression and function of CFTR at the apical membrane: NHERF, a protein known to anchor CFTR at the membrane and SLC26A14, an amino-acid transporter involved in cGMP-mediated F508del-CFTR channel activity.^{35,36}

Our study is the first report of human sweat proteome in pwCF.

Altogether, our results, based on a deep proteomic evaluation, provide evidence for an unbalanced proteolytic and proinflammatory activity of CF sweat, and alteration of signalling pathways involved in the integrity of the epidermal skin barrier. This CF-specific profile provides novel insight for AKP physiopathology and CFTR biology and may be used in the context of personalized medicine in CF.

AUTHOR CONTRIBUTIONS

Matthieu Cornet: Formal analysis (equal); Validation (equal); Visualization (equal); Writing – original draft (equal). **Thao Nguyen-Khoa:** Conceptualization (equal); Data curation (equal); Investigation (equal). **Mairead Kelly-Aubert:** Investigation (equal); Writing – review & editing (equal). **Vincent Jung:** Data curation (equal); Formal analysis (equal); Investigation (equal). **Frederique Chedevergne:** Data curation (equal); Investigation (equal). **Muriel Le Bourgeois:** Data

curation (equal); Investigation (equal). **Laura Aoust:** Data curation; Investigation (equal). **Kevin Roger:** Formal analysis; Visualization; Investigation (equal). **Chiara Guerrero:** Conceptualization (equal); Data curation; Investigation; Methodology; Project administration (equal); Resources (equal); Supervision (equal); Writing – original draft; Writing – review & editing (equal). **Isabelle Sermet-Gaudelus:** Conceptualization (equal); Funding acquisition (equal); Investigation; Methodology; Project administration (equal); Resources (equal); Supervision (equal); Writing – original draft; Writing – review & editing (equal).

ACKNOWLEDGEMENTS

This work was supported by the association Mucoviscidose ABCF2.

CONFLICT OF INTEREST

Isabelle Sermet-Gaudelus has been awarded a grant from Vertex Therapeutics (Vertex Innovation Award). She participates in Vertex Therapeutics, Eloxx Therapeutics and Proteostasis Therapeutics scientific boards.

DATA AVAILABILITY STATEMENT

Data available on request from the authors.

ETHICS STATEMENT

All participants signed an informed consent form for sweat collection and analysis (no. Eudract 2016-A00309-42; Comité de Protection des Personnes IDF2).

ORCID

Matthieu Cornet  <https://orcid.org/0000-0001-5466-9373>

REFERENCES

1. Elborn JS. Cystic fibrosis. *Lancet Lond Engl*. 2016;388(10059):2519–31. [https://doi.org/10.1016/s0140-6736\(16\)00576-6](https://doi.org/10.1016/s0140-6736(16)00576-6)
2. Alexopoulos A, Chouliaras G, Kakourou T, Dakoutrou M, Nasi L, Petrocheilou A, et al. Aquagenic wrinkling of the palms after brief immersion to water test as a screening tool for cystic fibrosis diagnosis. *J Eur Acad Dermatol Venereol J EADV*. 2021;35(8):1717–24. <https://doi.org/10.1111/jdv.17312>
3. Arkin LM, Flory JH, Shin DB, Gelfand JM, Treat JR, Allen J, et al. High prevalence of aquagenic wrinkling of the palms in patients with cystic fibrosis and association with measurable increases in transepidermal water loss. *Pediatr Dermatol*. 2012;29(5):560–6. <https://doi.org/10.1111/j.1525-1470.2011.01708.x>
4. Berk DR, Ciliberto HM, Sweet SC, Ferkol TW, Bayliss SJ. Aquagenic wrinkling of the palms in cystic fibrosis: comparison with controls and genotype-phenotype correlations. *Arch Dermatol*. 2009;145(11):1296–9. <https://doi.org/10.1001/archdermatol.2009.260>
5. Jacobi E, Solomon M, Avolio J, Shaw M, Gonska T, Ratjen F, et al. Aquagenic wrinkling of the palms in cystic fibrosis patients treated with ivacaftor. *J Cyst Fibros Off J Eur Cyst Fibros Soc*. 2022;21(2):e102–5. <https://doi.org/10.1016/j.jcf.2022.01.005>
6. Sermet-Gaudelus I, Nguyen-Khoa T, Hatton A, Hayes K, Pranke I. Sweat chloride testing and nasal potential difference (NPD) are primary outcome parameters in treatment with cystic fibrosis Transmembrane conductance regulator (CFTR) modulators. *J Pers Med*. 2021;11(8):729. <https://doi.org/10.3390/jpm11080729>
7. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol*. 2008;26(12):1367–72. <https://doi.org/10.1038/nbt.1511>
8. Nguyen-Khoa T, Hatton A, Drummond D, Aoust L, Schlatter J, Martin C, et al. Reclassifying inconclusive diagnosis for Cystic Fibrosis with new generation sweat test. *Eur Respir J*. 2022;60(2):2200209. <https://doi.org/10.1183/13993003.00209-2022>
9. Duguépéroux I, De Braekeleer M. The CFTR 3849+10kbC->T and 2789+5G->A alleles are associated with a mild CF phenotype. *Eur Respir J*. 2005;25(3):468–73. <https://doi.org/10.1183/09031936.05.10100004>
10. Yu Y, Prassas I, Muijtens CMJ, Diamandis EP. Proteomic and peptidomic analysis of human sweat with emphasis on proteolysis. *J Proteomics*. 2017;155:40–8. <https://doi.org/10.1016/j.jprot.2017.01.005>
11. Burat B, Reynaerts A, Baiwir D, Fleron M, Eppe G, Leal T, et al. Characterization of the human eccrine sweat proteome-A focus on the biological variability of individual sweat protein profiles. *Int J Mol Sci*. 2021;22(19):10871. <https://doi.org/10.3390/ijms221910871>
12. Csősz É, Emri G, Kalló G, Tsapralis G, Tozser J. Highly abundant defense proteins in human sweat as revealed by targeted proteomics and label-free quantification mass spectrometry. *J Eur Acad Dermatol Venereol J EADV*. 2015;29(10):2024–31. <https://doi.org/10.1111/jdv.13221>
13. Harshman SW, Pitsch RL, Smith ZK, O'Connor ML, Geier BA, Qualley AV, et al. The proteomic and metabolomic characterization of exercise-induced sweat for human performance monitoring: a pilot investigation. *PLoS One*. 2018;13(11):e0203133. <https://doi.org/10.1371/journal.pone.0203133>
14. Katchman BA, Zhu M, Blain Christen J, Anderson KS. Eccrine sweat as a biofluid for profiling immune biomarkers. *Proteomics Clin Appl*. 2018;12(6):e1800010. <https://doi.org/10.1002/prca.201800010>
15. Raiszadeh MM, Ross MM, Russo PS, Schaepper MA, Zhou W, Deng J, et al. Proteomic analysis of eccrine sweat: implications for the discovery of schizophrenia biomarker proteins. *J Proteome Res*. 2012;11(4):2127–39. <https://doi.org/10.1021/pr2007957>
16. Adewole OO, Erhabor GE, Adewole TO, Ojo AO, Oshokoya H, Wolfe LM, et al. Proteomic profiling of eccrine sweat reveals its potential as a diagnostic biofluid for active tuberculosis. *Proteomics Clin Appl*. 2016;10(5):547–53. <https://doi.org/10.1002/prca.201500071>
17. Baechle D, Flad T, Cansier A, Steffen H, Schitteck B, Tolson J, et al. Cathepsin D is present in human eccrine sweat and involved in the postsecretory processing of the antimicrobial peptide DCD-1L. *J Biol Chem*. 2006;281(9):5406–15. <https://doi.org/10.1074/jbc.m504670200>
18. Murakami M, Ohtake T, Dorschner RA, Gallo RL, Schitteck B, Garbe C. Cathelicidin anti-microbial peptide expression in sweat, an innate defense system for the skin. *J Invest Dermatol*. 2002;119(5):1090–5. <https://doi.org/10.1046/j.1523-1747.2002.19507.x>
19. Park JH, Park GT, Cho IH, Sim SM, Yang JM, Lee DY. An antimicrobial protein, lactoferrin exists in the sweat: proteomic analysis of sweat. *Exp Dermatol*. 2011;20(4):369–71. <https://doi.org/10.1111/j.1600-0625.2010.01218.x>
20. Schitteck B, Hipfel R, Sauer B, Bauer J, Kalbacher H, Stevanovic S, et al. Dermcidin: a novel human antibiotic peptide secreted by sweat glands. *Nat Immunol*. 2001;2(12):1133–7. <https://doi.org/10.1038/ni732>
21. Moyer J, Wilson D, Finkelshtein I, Wong B, Potts R. Correlation between sweat glucose and blood glucose in subjects with

- diabetes. *Diabetes Technol Ther.* 2012;14(5):398–402. <https://doi.org/10.1089/dia.2011.0262>
22. Baker LB. Physiology of sweat gland function: the roles of sweating and sweat composition in human health. *Temp Austin Tex.* 2019;6(3):211–59. <https://doi.org/10.1080/23328940.2019.1632145>
 23. Kasperek P, Ileninova Z, Zbodakova O, Kanchev I, Benada O, Chalupsky K, et al. KLK5 and KLK7 ablation fully rescues lethality of Netherton syndrome-like phenotype. *PLoS Genet.* 2017;13(1):e1006566. <https://doi.org/10.1371/journal.pgen.1006566>
 24. Dallacasagrande V, Hajjar KA. Annexin A2 in inflammation and host defense. *Cells.* 2020;9(6):E1499. <https://doi.org/10.3390/cells9061499>
 25. Kwak EJ, Hong JY, Kim MN, Kim SY, Kim SH, Park CO, et al. Chitinase 3-like 1 drives allergic skin inflammation via Th2 immunity and M2 macrophage activation. *Clin Exp Allergy J Br Soc Allergy Clin Immunol.* 2019;49(11):1464–74. <https://doi.org/10.1111/cea.13478>
 26. Leonardi S, Parisi GF, Capizzi A, Manti S, Cuppari C, Scuderi MG, et al. YKL-40 as marker of severe lung disease in cystic fibrosis patients. *J Cyst Fibros Off J Eur Cyst Fibros Soc.* 2016;15(5):583–6. <https://doi.org/10.1016/j.jcf.2015.12.020>
 27. Burns K, Clatworthy J, Martin L, Martinon F, Plumpton C, Maschera B, et al. Tollip, a new component of the IL-1RI pathway, links IRAK to the IL-1 receptor. *Nat Cell Biol.* 2000;2(6):346–51. <https://doi.org/10.1038/35014038>
 28. Cheng KC, Lin RJ, Cheng JY, Wang SH, Yu JC, Wu JC, et al. FAM129B, an antioxidative protein, reduces chemosensitivity by competing with Nrf2 for Keap1 binding. *EBioMedicine.* 2019;45:25–38. <https://doi.org/10.1016/j.ebiom.2019.06.022>
 29. Bonnart C, Deraison C, Lacroix M, Uchida Y, Besson C, Robin A, et al. Elastase 2 is expressed in human and mouse epidermis and impairs skin barrier function in Netherton syndrome through filaggrin and lipid misprocessing. *J Clin Invest.* 2010;120(3):871–82. <https://doi.org/10.1172/jci41440>
 30. Sandilands A, Sutherland C, Irvine AD, McLean WHI. Filaggrin in the frontline: role in skin barrier function and disease. *J Cell Sci.* 2009;122(Pt 9):1285–94. <https://doi.org/10.1242/jcs.033969>
 31. Palmer CNA, Irvine AD, Terron-Kwiatkowski A, Zhao Y, Liao H, Lee SP, et al. Common loss-of-function variants of the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis. *Nat Genet.* 2006;38(4):441–6.
 32. Weidinger S, Illig T, Baurecht H, Irvine AD, Rodriguez E, Diaz-Lacava A, et al. Loss-of-function variations within the filaggrin gene predispose for atopic dermatitis with allergic sensitizations. *J Allergy Clin Immunol.* 2006;118(1):214–9. <https://doi.org/10.1016/j.jaci.2006.05.004>
 33. Bonn  S, Gilbert B, Hatzfeld M, Chen X, Green KJ, van Roy F. Defining desmosomal plakophilin-3 interactions. *J Cell Biol.* 2003;161(2):403–16. <https://doi.org/10.1083/jcb.200303036>
 34. Rongioletti F, Tomasini C, Crovato F, Marchesi L. Aquagenic (pseudo) keratoderma: a clinical series with new pathological insights. *Br J Dermatol.* 2012;167(3):575–82. <https://doi.org/10.1111/j.1365-2133.2012.11003.x>
 35. Seidler U, Singh AK, Cinar A, Chen M, Hillesheim J, Hogema B, et al. The role of the NHERF family of PDZ scaffolding proteins in the regulation of salt and water transport. *Ann N Y Acad Sci.* 2009;1165(1):249–60. <https://doi.org/10.1111/j.1749-6632.2009.04046.x>
 36. Ruffin M, Mercier J, Calmel C, Mesinele J, Bigot J, Sutanto EN, et al. Update on SLC6A14 in lung and gastrointestinal physiology and pathophysiology: focus on cystic fibrosis. *Cell Mol Life Sci CMLS.* 2020;77(17):3311–23. <https://doi.org/10.1007/s00018-020-03487-x>

How to cite this article: Cornet M, Nguyen-Khoa T, Kelly-Aubert M, Jung V, Chedevergne F, Le Bourgeois M, et al. Proteomic profiling of sweat in patients with cystic fibrosis provides new insights into epidermal homeostasis. *Skin Health Dis.* 2022;e161. <https://doi.org/10.1002/ski2.161>

Representation and quantification Of Module Activity from omics data with rROMA

Matthieu Cornet^{1,2,3,#}, Matthieu Najm^{1,2,3,#}, Luca Albergante^{1,2,3}, Andrei Zinovyev^{1,2,3}, Isabelle Sermet-Gaudelus^{4,5,6}, Véronique Stoven^{1,2,3}, Laurence Calzone^{1,2,3} and Loredana Martignetti^{1,2,3 *}

¹ INSERM U900, 75428 Paris, France, ² Center for Computational Biology, Mines ParisTech, PSL Research University, 75006 Paris, France, ³ Institut Curie, PSL Research University, 75248 Paris, France, ⁴ Faculté de Médecine, Université de Paris, Paris, France, ⁵ Institut Necker Enfants Malades, INSERM U1151, Paris, France, ⁶ AP-HP. Centre - Université Paris Cité; Hôpital Necker Enfants Malades, Centre de Référence Maladie Rare - Mucoviscidose, Paris, France.

These authors contributed equally to this work; * To whom correspondence should be addressed.

Abstract

In many analyses of high-throughput data in systems biology, calculating the activity of a set of genes rather than focusing on the differential expression of individual genes has proven to be efficient and informative. Here, we present the rROMA software package for fast and accurate computation of the activity of gene sets with coordinated expression. We applied rROMA to cystic fibrosis, highlighting biological mechanisms potentially involved in the establishment and progression of the disease and the associated genes. Source code and documentation are available at <https://github.com/sysbio-curie/rROMA>.

Contact: loredana.martignetti@curie.fr

1 Introduction

A signaling pathway can be defined as a set of genes or proteins that interact to pass on information from the exterior to the interior of a cell. In many diseases, some pathways are altered and the deregulations may come from one or several genes in this pathway, which may differ from one patient to another. Quantification of pathway activity via numerical scores using high throughput measurement of gene and protein expression is widely applied to transform the gene-level data into interpretable gene sets that can reveal biological heterogeneity across samples.

Here we present rROMA, a user-friendly and interactive R implementation of the ROMA algorithm that quantifies the activity of a gene set (or module) by its first weighted principal component (PC) (Martignetti et al, 2016). Here, gene sets or modules referred to canonical pathways annotated by domain experts or derived by external databases. In ROMA, the co-variance across samples of the genes composing the module is interpreted as the result of the action of a hidden factor on the expression of the genes. This setting corresponds to the simplest uni-factor linear model of gene expression regulation (Schreiber and Baumann, 2007).

The rROMA package provides novel functionalities for the analysis, visualization and reporting of active/inactive modules. It allows to highlight two possible configurations of the expression of the genes of a module under the effect of a regulatory factor: we called *shifted* module the case where the genes of the module are collectively displaced to one side with respect to the center of the distribution (Fig. 1A), while *overdispersed* module is the case where the amount of variance explained by the first PC of a given gene set is significantly higher than expected in the global gene expression distribution and the genes are found dispersed on both sides of the center of the global gene expression distribution (Fig. 1B).

Moreover, the algorithm implements different ways to identify outlier genes that can significantly influence the results. Finally, several functions for differential analysis and graphical visualization of the results are provided.

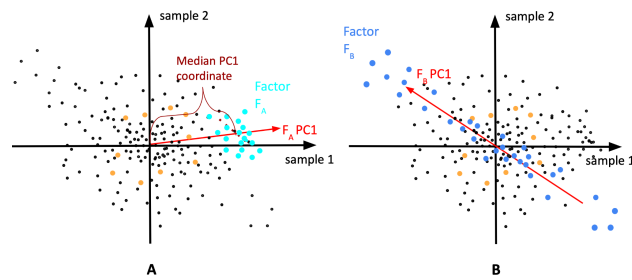


Fig. 1. Representation of genes in the case of two samples. Each dot represents one gene, its horizontal (resp. vertical) value corresponding to its expression in sample 1 (resp. sample 2). Genes associated with Factor A are plotted in light blue and the corresponding PC1 direction is plotted in red (A). This represents a shifted pathway, as assessed by a median of gene projections onto PC1 direction far from the origin of the distribution. Genes associated with Factor B are plotted in dark blue (B) and the corresponding PC1 direction is plotted in red. This represents an overdispersed pathway, as the PC1 is well aligned with the dots' distribution. Larger dots correspond to the genes with the highest score, highlighting their importance for the pathway. Genes in yellow are neither overdispersed nor shifted, as PC1 explains a relatively small fraction of variance (not represented on the figure) and the median of projections onto PC1 is close to the origin for this group of genes.

2 Implementation

rROMA is distributed as an open-source R software package and is available on GitHub: www.github.com/sysbio-curie/rROMA. A detailed vignette to reproduce all the analyses presented in this paper is also available.

3 Results

We applied rROMA to investigate the activity of pathways in airway epithelial cells from cystic fibrosis (CF) patients and from healthy donors. More precisely, we compared the transcriptomes of primary cultures of airway epithelial cells from patients (N=6) with those of healthy controls (N=6), based on RNAseq data publicly available in the NCBI's GEO database, under the accession ID GSE 176121 (Rehman et al, 2021).

Here, the Molecular Signature Database MsigDB hallmark gene set collection (Liberzon et al, 2015) was selected to ease interpretation. However, to provide a more complete view of the biological processes involved in a study, rRoma can be applied using multiple pathway databases. rRoma was run by specifying the pathway database to use and the expression matrix to analyze, as shown in the accompanying vignette.

Once rRoma has finished running the analysis, activated pathways can be determined by looking at the *ModuleMatrix* output. Pathways with a *ppv Median Exp* lower than a certain threshold were deemed as *shifted*, while those with a *ppv L1* lower than this threshold were *overdispersed*. Pathway activity across samples can be plotted with the *Plot.Genesets.Samples* function, and the top contributing genes in each pathway are determined by visualizing gene weights with the function *PlotGeneWeight*. Boxplot of the activity scores based on predefined groups can also be plotted for differential analysis. In this vignette, all highlighted pathways behaved significantly differently in CF patients versus healthy donors.

In our example, out of the 50 hallmark pathways tested, 3 were significantly active: Fatty acid metabolism, apical surface, and coagulation. The *Fatty acid metabolism* pathway has significantly different activity scores between CF patients and healthy donors. In the context of CF, this pathway has been extensively studied, and its deregulation is a well known CF phenotype (for a review, see Strandvik, 2010). This illustrates the ability of rROMA to retrieve dysregulations from the transcriptomic data.

The *apical_surface* pathway is more difficult to interpret. This might actually highlight differences arising during cell differentiation, and it could thus be related to cell culture rather than to the disease. Finally, the *coagulation* pathway, the only overdispersed pathway, seems to be linked to one specific gene with a very high associated weight: *gelsolin* (*GSN*). We observe that *GSN* is by far the top contributing gene to the activity score of the *COAGULATION* pathway. *Gelsolin* has been reported as playing a role for CFTR activation (Cantiello 2001, Vasconcellos et al 1994). Overall, in this case study, rRoma highlighted a relevant mechanism in the context of CF, a potential bias due to cell culture, and an interesting gene which could be further investigated.

Many hyperparameters can be specified and changed to modify rRoma speed, precision, or behavior regarding outliers. Details about all available hyperparameters are described in the vignette. The computational time required to run the algorithm typically depends on the number of studied pathways and their relative size. It also depends on whether parallelization is enabled. Regarding the example discussed here, the algorithm ran in approximately 3 minutes and 15 seconds on a MacBook Pro equipped with a 2,6 GHz Intel Core i7 6 cores processor. A single 60 genes pathway took roughly 5 seconds to be analyzed. Note that parallelization was not used but would have increased the speed of the analysis if used.

In summary, this work indicates that rROMA is capable of identifying genetic pathways contributing to disease-associated transcriptomics enabling a clearer interpretation of results from a biological point of view, which allows interpreting cellular changes in a more holistic and functional way.

Acknowledgements

The authors would like to thank Pierre Gestraud for insightful discussions about the R implementation.

Funding

This work was supported by the European Commission's Horizon 2020 Program, H2020-SC1-DTH-2018-1, "iPC individualizedPaediatricCure" (ref. 826121) and by the French government under management of Agence Nationale de la Recherche as part of the "Investissements d'avenir" program, reference ANR-19-P3IA-0001 (PRAIRIE 3IA Institute). MC was supported by Fondation pour la Recherche Médicale (FRM), and MN was supported by Vaincre La Mucoviscidose (VLM) grant number 20190502488.

Conflict of Interest: none declared.

References

- Cantiello H, (2001) Role of actin filament organization in CFTR activation. *Pflügers Archiv* **443**,S75–S80
- Favia M. *et al*, (2019) An Intriguing Involvement of Mitochondria in Cystic Fibrosis. *J Clin Med.*, **6**;8(11):1890
- Liberzon A. *et al*. (2015) The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.*, **23**;1(6):417-425
- Martignetti L. *et al*. (2016) ROMA: Representation and Quantification of Module Activity from Target Expression Data. *Front Genet.*, **19**;7:18
- Rehman T *et al*. (2021) Inflammatory cytokines TNF- α and IL-17 enhance the efficacy of cystic fibrosis transmembrane conductance regulator modulators. *J Clin Invest* **16**;131(16)
- Schreiber A.W. and Baumann U. (2007) A framework for gene expression analysis. *Bioinformatics*, **15**;23(2):191-7
- Strandvik B. (2010) Fatty acid metabolism in cystic fibrosis. *Prostaglandins Leukot Essent Fatty Acids* **83**(3):121-9
- Vasconcellos C.A. *et al*, (1994) Reduction in viscosity of cystic fibrosis sputum in vitro by gelsolin. *Science*, **18**;263(5149):969-71

RÉSUMÉ

La mucoviscidose est une maladie provoquée par une mutation d'un gène produisant une protéine dont l'absence ou le dysfonctionnement entraîne des lésions au niveau des poumons et un déclin des fonctions respiratoires. Le pronostic des patients porteurs de la mutation la plus courante a récemment été grandement amélioré grâce à la mise au point de nouvelles thérapies. Cependant, la réponse à ces traitements n'est pas homogène parmi les patients, et leur mécanisme d'action n'est pas entièrement compris. En particulier, nous ne disposons pas de marqueur biologique ou clinique qui fasse consensus pour mesurer l'effet d'un traitement. L'objectif de cette thèse a été de caractériser les réponses cliniques à ces thérapies, en particulier chez les jeunes patients, ainsi que de mieux comprendre leurs mécanismes d'action au niveau cellulaire.

MOTS CLÉS

approches omiques, bioinformatique, mucoviscidose, étude clinique

ABSTRACT

Cystic fibrosis is a disease caused by a mutation in a gene that produces a protein whose absence or malfunction leads to lung damage and a decline in respiratory function. The prognosis for patients with the most common mutation has recently been greatly improved by the development of new therapies. However, the response to these therapies is not homogeneous among patients, and their mechanism of action is not fully understood. In particular, we do not have a universally accepted biological or clinical marker to measure the effect of a treatment. The objective of this thesis was to characterize the clinical responses to these therapies, particularly in young patients, and to better understand their mechanisms of action at the cellular level.

KEYWORDS

Omics approaches, bioinformatics, cystic fibrosis, clinical study