

**FICHE SUJET DE THESE**

Sujet N° (à remplir par l'ED) :	<b>FINANCEMENT :</b> <input checked="" type="checkbox"/> Demandé <input type="checkbox"/> Acquis	Origine du financement : 100% CDE
Titre de la thèse : <b>Prédiction de l'efficacité des thérapies ciblées chez des patients atteints de pathologies auto-immunes. Elaboration de nouvelles approches algorithmiques de données OMIQUES.</b>		3 mots-clés : Maladies auto-immunes, Extraction de règles, Réseaux de convolution
Unité/équipe encadrante : <b>Lymphocytes B et Auto-Immunité (LBAI, UMR 1227)</b>		
Directeur de thèse : Christophe Jamin (UMR1227), directeur et Pascal Redou (UMR1101), co-directeur		N° de tél : 02 98 22 33 84 Mail : christophe.jamin@univ-brest.fr
<p><u>Contexte socioéconomique et scientifique (env. 10 lignes) :</u>  Les maladies auto-immunes systémiques sont un groupe de maladies inflammatoires chroniques dont les diagnostics sont difficiles à établir et les traitements aléatoires. Leur caractéristique est la présence d'auto-anticorps dans le sérum (dirigés contre des composants du Soi). Alors que chaque entité clinique peut être considérée comme « rare », l'ensemble constitue la troisième cause de morbidité dans le monde. Ces maladies sont au centre du projet Européen IMI PRECISESADS pour lequel le LBAI (UMR1227) est partenaire. Le but est d'utiliser la puissance des techniques "OMIQUES" (génomique, transcriptomique, épigénomique, métabolomique, protéomique), et la bioinformatique pour les re-classifier. Cette approche n'a jamais été appliquée à l'échelle du patient, le privant des bénéfices de nouveaux agents thérapeutiques qui ne peuvent lui être proposés en raison des classifications cliniques actuelles. Notre hypothèse est que l'identification de signatures moléculaires spécifiques chez les patients atteints de maladies auto-immunes systémiques permettra aux cliniciens d'adapter leurs traitements.</p> <p>L'objectif de cette thèse sera d'extraire des règles de classification portant sur les millions de variables renseignées pour chaque patient afin d'identifier les marqueurs biologiques capables de prédire l'efficacité d'une thérapie ciblée (développement de tests compagnons). La thèse s'inscrit dans un projet impliquant des chercheurs du LBAI et du Laboratoire de Traitement de l'Information Médicale (LaTIM, UMR1101), dans le but d'aller bien au-delà de l'analyse statistique classique ou de l'apprentissage automatique de type « boîte noire ». Notre résultat préliminaire prometteur consiste en une transformation judicieuse en images de données patients/variables quelconques, afin de leur appliquer des réseaux de convolution (apprentissage profond) puis d'en extraire des patterns permettant la classification et l'extraction de règles sur les variables. Nous développerons cette méthode, en l'optimisant via l'analyse mathématique et la détermination d'un réseau de convolution ad hoc, et en l'appliquant sur les données "OMIQUES".</p>		
<p><u>Hypothèses et questions posées (env. 8 lignes) :</u>  Notre hypothèse initiale repose sur l'efficacité de la méthodologie consistant à enchaîner l'algorithme évoqué ci-dessus de transformation des données en images (pixellisation) avec le traitement de ces images au moyen de réseaux convolutionnels. Notre hypothèse est déjà étayée par de nombreux tests préliminaires concluants effectués sur des données publiques (kaggle). Cependant, cette hypothèse soulève plusieurs questions. Quelle stratégie faut-il développer pour une pixellisation optimale des données ? La représentation bi-dimensionnelle peut-elle être généralisée ? Existe-t-il une stratégie générique optimale quelles que soient les données ou bien faut-il des stratégies ad hoc qui soient adaptées à chaque type de données traitées ? Par ailleurs, pour être exploitable à grande échelle en vue de la prédiction de l'efficacité thérapeutique d'un traitement, un test compagnon doit s'appuyer sur un nombre limité de données ; ce qui va à l'opposé de ces méthodes convolutionnelles, qui ont tendance à s'étaler sur une large zone de données. Quelles règles simples, s'appuyant sur le plus petit nombre de mesures, permettront d'extraire des résultats de l'apprentissage, et comment les extraire ? Une mise en commun des compétences biologiques (UMR1227) et informatiques (UMR1101) est indispensable pour la réussite du projet.</p>		
<p><u>Grandes étapes de la thèse (env. 12 lignes) :</u>  La première étape de la thèse sera de comprendre l'efficacité de la méthode. Notre hypothèse est que cette méthode fonctionne parce qu'elle permet, avant l'étape d'apprentissage profond, de rapprocher les données liées entre elles, soit du fait de l'expertise de l'immunologiste, soit par mesure de corrélations. Cette hypothèse sera à valider soit en étudiant le fonctionnement interne de la méthode, soit expérimentalement.</p> <p>La deuxième étape consistera à s'appuyer sur la réponse à la première pour produire un algorithme dédié à notre objectif. Il n'y a en effet aucune raison pour que l'organisation des données "OMIQUES" en grille de pixels bidimensionnelle soit optimale. Concernant la partie relevant de l'apprentissage profond, il faudra également identifier les réseaux convolutionnels les plus efficaces. La dernière étape, constituant l'objectif ultime de la thèse, nous permettra d'obtenir une méthode d'extraction de règles, s'appuyant sur la méthode d'apprentissage validée mais produisant des règles qui utilisent le plus petit nombre de mesures. L'application envisagée est en effet la production de « tests compagnons » pour des traitements, c'est à dire de tests spécifiques à chaque traitement, et donnant une prédiction de leur efficacité sur une cible particulière. Pour être utilisables systématiquement, ces tests devront être constitués de données génomiques ou cytométriques partielles. Il faudra donc identifier, pour chaque test, quelles sont les quelques données suffisantes à résumer tout le réseau convolutionnel.</p>		
<p><u>Compétences scientifiques et techniques requises par le candidat (2 lignes) :</u>  Le profil recherché est celui d'un (bio-)informaticien ou mathématicien appliqué ayant une bonne connaissance, y compris sur les aspects mathématiques, de l'analyse de données et de l'apprentissage automatique, qu'il soit profond ou non.</p>		
<p><u>3 publications de l'équipe d'accueil relatives au domaine (5 dernières années) :</u>  -Jamin C, Le Lann L, Alvarez-Errico D, Barbaroja N, Cantarert T, Ducreux J, Dufour AM, Gerl V, Kniesch K, Neves E, Trombetta E, Alarcón-Riquelme M, Marañon C, Pers JO. Multi-center harmonization of flow cytometers in the context of the European "PRECISESADS" project. Autoimmun Rev. 2016 Nov;15(11):1038-1045.  -Foulquier N, Redou P, Le Gal C, Rouvière B, Pers JO, Saraux A. Pathogenesis-based treatments in primary Sjogren's syndrome using artificial intelligence and advanced machine learning techniques: a systematic literature review. Hum Vaccin Immunother. 2018;14(11):2553-2558.  -Foulquier N, Le Gal C, Redou P, Pers JO. Topological data representation for automatic extraction of association rules using common patterns, soumis à Journal of Artificial Intelligence Research.</p>		
<p><u>Collaborations nationales et internationales :</u>  -Projet collaboratif UMR1227-UMR1101.  -Projet européen Innovative Medicines Initiative (IMI) "Molecular Reclassification to Find Clinically Useful Biomarkers for Systemic Autoimmune Diseases (PRECISESADS)"  -Projet européen Horizon 2020 Research and Innovation Programme "HARMONization and integrative analysis of regional, national and international Cohorts on primary Sjögren's Syndrome (pSS) towards improved stratification, treatment and health policy making disease."  -Projet européen Innovative Medicines Initiative (IMI) "NEw Clinical Endpoints in primary Sjögren's Syndrome: an Interventional Trial based on stratifying patients (NECESSITY)"</p>		