

THESIS TOPIC

Subject N° (to be completed by the ED):	FUNDING: <input checked="" type="checkbox"/> Requested <input type="checkbox"/> Acquired	Funding origin: CDE 100%
Thesis title: Prediction of the efficacy of targeted therapies in patients suffering autoimmune diseases. Elaboration of new algorithmic approaches of OMIC data.		3 keywords: Autoimmune diseases, Extraction of rules, Convolutional networks
Unit / team: Lymphocytes B et Auto-Immunité (LBAI, UMR 1227)		
Supervisor's name: Christophe Jamin (UMR1227), directeur et Pascal Redou (UMR1101), co-supervisor		Phone number: +332 98 22 33 84 Email address: christophe.jamin@univ-brest.fr
<p>Socio-economic and scientific context (approximately 10 lines): Systemic autoimmune diseases are a group of chronic inflammatory diseases whose diagnoses are difficult to establish and treatments uncertain. Their common feature is the presence of autoantibodies in the serum (directed against components of the Self). While each clinical entity can be considered as "rare", the whole is the third leading cause of morbidity in the world. These diseases are at the center of the European IMI PRECISESADS project for which LBAI (UMR1227) is a partner. The goal of this project is to use the power of the so-called "OMICS" techniques (genomics, transcriptomics, epigenomics, metabolomics, proteomics), and bioinformatics to re-classify the diseases. This "OMIC" approach has never been applied at the patient level, depriving it of the potential benefits of new therapeutic agents that can not be offered because of current clinical classifications. Our main hypothesis is that the identification of specific molecular signatures in patients with systemic autoimmune diseases will allow clinicians to adapt their treatments. The objective of this thesis will be to extract classification rules concerning the millions of variables that are reported for each patient in the PRECISESADS project in order to identify the biological markers that can predict the efficacy of a targeted therapy (development of companion tests). The thesis is part of a study involving researchers from LBAI and the Laboratoire de Traitement de l'Information Médicale (LaTIM, UMR1101), with the aim of going beyond classical statistical analysis or automatic learning such as "black box" environment. Our most promising preliminary result is a judicious transformation of any patient variable data into images, in order to apply convolutional networks (deep learning) to them and then to extract patterns allowing classification and extraction of rules on variables. We will develop this method, by optimizing it thanks to the mathematical analysis and the determination of an ad hoc convolution network, and applying it on the "OMIC" data.</p>		
<p>Working hypothesis and aims (approximately 8 lines): Our initial hypothesis is based on the effectiveness of the methodology of linking the aforementioned algorithm of transformation of data into images (pixelation) with the processing of these images using convolutional networks. Our hypothesis is already supported by numerous conclusive preliminary tests carried out on public data (kaggle). However, this hypothesis raises several questions. What strategy should be developed for optimal pixelation of data? Can two-dimensional representation be generalized? Is there an optimal generic strategy regardless of the data or do we need ad hoc strategies that are appropriate for each type of data processed? Moreover, to be achievable on a large scale for the purpose of predicting the therapeutic efficacy of a treatment, a companion test must be based on a limited number of data; which is the opposite of these convolutional methods, which tend to spread over a large data area. What simple rules, based on the smallest number of measures, will allow the extraction of learning outcomes, and how to extract them? A pooling of biological (UMR1227) and informatical (UMR1101) skills is mandatory for a successful achievement of the project.</p>		
<p>Main milestones of the thesis (approximately 12 lines): The first step of the thesis will be to understand the effectiveness of the method. Our hypothesis is that this method works because it allows, before the deep learning step, to reconcile the data linked together, either because of the immunologist's expertise or because of evaluation of correlations. This hypothesis will be validated either by studying the internal functioning of the method, or experimentally. The second step will be to rely on the answer emerging from the first step to produce an algorithm dedicated to our goal. Indeed, there is no reason for the organization of "OMIC" data in two-dimensional pixel grid to be optimal. Regarding the deep learning aspect, it will also be necessary to identify the most efficient convolutional networks. The last step, constituting the ultimate goal of the thesis, will allow us to obtain a rule extraction method, based on the validated learning method but producing rules that use the smallest number of measures. The intended application is indeed the production of "companion tests" for treatments, i.e. tests specific to each treatment, and giving a prediction of their efficacy on a particular target. To be used systematically, these tests must be made of partial genomic or cytometric data. It will therefore be necessary to identify, for each test, what are the few sufficient data to summarize the entire convolutional network.</p>		
<p>Scientific and technical skills required by the candidate (2 lines): The desired profile is a (bio-)informatician or an applied mathematician with good knowledges, including mathematical aspects, of data analysis and machine learning, whether deep or not.</p>		
<p>3 publications from the team related to the topic (last 5 years): -Jamin C, Le Lann L, Alvarez-Errico D, Barbarroja N, Cantaert T, Ducreux J, Dufour AM, Gerl V, Kniesch K, Neves E, Trombetta E, Alarcón-Riquelme M, Marañón C, Pers JO. Multi-center harmonization of flow cytometers in the context of the European "PRECISESADS" project. <i>Autoimmun Rev.</i> 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. <i>Hum Vaccin Immunother.</i> 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 <i>Journal of Artificial Intelligence Research</i>.</p>		
<p>National and international collaborations: -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>		