Séquençage à haut débit des récepteurs antigéniques dans les hémopathies lymphoïdes

Mots clés :
- Directeur de thèse : Martin Weigt
- Co-encadrant(s) :
- Unité de recherche : Laboratoire de Biologie Computationnelle et Quantitative
- Ecole doctorale : École Doctorale Informatique, Télécommunications, Electronique de Paris
- Domaine scientifique principal : Sciences et technologies de l'information et de la communication

Résumé du projet de recherche (Langue 1)

B cells constitute a major cellular component of the adaptive immune system. They directly bind to pathogens through a specific receptor (B-cell receptor, BCR) and secrete antibodies, the soluble form of the BCR. Each B cell expresses a unique BCR which allows the recognition of a particular set of molecular structure (antigen). The immune system has the ability to respond to almost any antigen to which it is exposed because of the incredible diversity of receptor, about 10e12 molecules. This diversity originates from complex genetic mechanisms during BCR assembly and maturation. Analyses of the BCR sequences have been instrumental in studying the immune system of healthy individuals, but also in various fields of medicine such as infections, allergies, cancer and auto-immunity. In cancer, BCR sequences analysis is regularly used for monitoring B cell tumors, as they provide critical diagnostic, prognostic and response to treatment information. In this project, we shall develop new pipelines to analyze these complex data and produce biologically interpretable results in order to analysis BCR repertoire, particularly in the field of B cell tumors. As a first step, we are carrying out a comparative study and a rigorous evaluation of the existent tools to identify their limitations. We have already identified some of them and developed some methods to address these problems, mainly in terms of visualization and sequence comparisons (manuscripts in preparation). However, a number of other improvements and functionalities need to be implemented to provide a more robust and complete set of tools. They include repertoire diversity estimation, affinity maturation analysis, BCR clonal architecture, 3D antigen-biding region modeling, large-scale inter-patient comparisons, detection of rare tumor-specific among vast mix of heterogeneous BCR sequences, etc. To address these points, the computational methods must be fast and efficient since they have to deal with millions of sequences per individual. For that, we shall explore big-data algorithms, especially those employing data mining and machine learning methods that have been poorly explored until now in this domain. To process the vast amounts of data we could use MapReduce strategy, a style of parallel computing that has been implemented in several systems, including Google and hadoop.

Résumé du projet de recherche (Langue 2)

A challenge in developing computational pipelines for B-cell receptor repertoire sequencing analysis consists in the huge amount of data to be treated. It necessitates efficient algorithms combined with parallel programming and distributed processing approaches. Another important point is to deal with real data coming NGS experiments, possibly containing sequence errors, missing parts, and inconsistencies. Here, we shall explore corrective strategies to identify patterns/reasons for missing/errors and recode data correctly. Although several computational tools have been previously published, most of them are only used by highly specialized researchers. As NGS is entering medical diagnostic and BCR sequence analysis essential for B cell tumors management, our aim is to produce a set of user-friendly tools available for the medical community. Thus, we will work in close collaboration with Pr. Frédéric Davi’s team (Department of Hematology, Hôpital Pitié-Salpêtrière), who has data available for large series of well-characterized patients. These will be used to validate and improve our pipeline. We will also collaborate with Mathieu Giraud’s team at Laboratoire d’Informatique Fondamentale de Lille, who developed a software and web-server for antigen receptor repertoire analysis. Ours approaches are complementary and indeed, this collaboration will yield better results and improved tools.

Informations complémentaires (Langue 1)

This project also involves international collaborations with: (i) Dr. Kostas Stamatopoulos’s team at Institute of Applied Biosciences, CERTH, Thessaloniki, Greece, who has a long-standing experience in the field of immunogenetics; (ii) the EuroClonality consortium, a network of European diagnostic laboratories working on standardization of NGS methods in the field tumors of the immune system. The work of the Ph.D. candidate will benefit from these collaborations through participation to joint scientific meetings, conferences and possible visits to associated countries.
References