M2 internship: Applied mathematics for biological data processing and analysis

I. Background

Modern high-throughput "omic" technologies have high potential for deciphering the organism functions and developing personalized diagnostic and therapeutic strategies. In particular, metabolomics by liquid chromatography coupled to high-resolution mass spectrometry (LC-HRMS) provides non-targeted, highly sensitive detection of the content from a biological or clinical sample which, in turns, enables refined phenotype characterization of the individual.
Annotation (i.e. determining the molecule structure from the measured mass, retention time, fragmentation pattern) is a major challenge due to the high chemical variability of metabolites. Current software tools for MS data preprocessing (peak detection, grouping of isotopes, adducts, fragments) like XCMS and CAMERA provide partial results which must be manually checked by the expert chemist. In addition, public database query from a single mass result in no or multiple candidate(s) which must also be filtered individually.
Within the French Metabolomics and Fluxomics Infrastructure (MetaboHUB, www.metabohub.fr/en), CEA (www.cea.fr/english-portal) has strong expertise in MS data processing and analysis at the LADIS laboratory (Technological Research Division) together with analytical acquisition and MS and MS/MS spectra annotation at the LEMM laboratory (Life Sciences Division).

II. Project

The Master 2 candidate will develop innovative algorithms for refined LC-HRMS metabolomics annotation: chemical formula, structure, metabolic network. First, a precise peak detection within the raw data (MS and MS^2) will be implemented. Second, chemical rules (isotopes, fragments, adducts) will be defined with the experimental partners and integrated within the software. Third, public database query and their visualization via graphs will provide biological and chemical links between the annotated molecules. Fourth, bayesian analysis will be used to sort candidates in case of multiple annotation of the same peak.
Data will be provided by the experimental partners from MetaboHUB. The developed software tools will be implemented within the bioinformatics resource (http://workflow4metabolomics.org) jointly developed and maintained by the French Bioinformatics Institute and MetaboHUB.

Key words: applied mathematics, signal processing, statistics, mass spectrometry, bioinformatics

III. Profile

The successful candidate will be graduated in applied mathematics (including signal processing and statistics) and have strong expertise in the R and C languages. The candidate will have specific support in the non-mathematical aspects of the project (e.g. chemistry). The internship is intended to be followed by a PhD.

IV. Application (deadline November 30th, 2014)

Please send your CV, letter of motivation, and contact information of two reference supervisors to:
Etienne Thévenot (etienne.thevenot@cea.fr)
Laboratory for Data Analysis and Smart Systems
CEA, Saclay Center, build. 565 (DRT/LIST/DM2I/LADIS)
F-91191 Gif-sur-Yvette Cedex, France
Website: http://www-centre-saclay.cea.fr/en, www-list.cea.fr/en