Institut Pasteur
25 rue du Docteur Roux
75015 Paris
France

December, the 7th 2017

Title: Predicting disease heterogeneity from genetic data

Context & objectives

Applications are invited for an internship at the Institut Pasteur within the Statistical Genetics group in the Center of Bioinformatics, Biostatistics and Integrative Biology (C3BI).

Several common human diseases are not single entity but rather composed of multiple heterogeneous disease subtypes. For example inflammatory bowel disease encompasses Crohn's disease and ulcerative colitis, while breast cancer can be either ER (estrogen receptor) positive or negative. Subtypes are expected for many other diseases, but have not yet been characterized. Identifying latent disease subtypes is important both for elucidating the causes of disease, and for effective clinical treatment. Genetic and genomic data, which have led to new insights about causal variants, genes, pathways, and cell types for disease, have the potential to be informative also about disease subtypes. Indeed, disease heterogeneity commonly implies heterogeneous genetic architecture and subtype's specific pattern of gene expression and other endophenotypes. Nevertheless, despite the topic is of broad interest for the genetic community, little progress has been achieved in developing the tools for assessing and quantifying potential disease heterogeneity from genetic data.

The internship involves statistical method application/development and the optimization of their implementation. The objectives of the candidate will be multiple folds:

1) Understanding the basis of genetic-phenotypes associations and familiarizing with genetics and genomics data.
2) Being able to summarize theoretical models leading to genetic disease heterogeneity.
3) Exploring the relative performances of statistical methods for identifying disease subtypes in simulated data. Methods will include in particular principal components analysis (PCA) and other singular-value decomposition approaches applied to genotype matrix.
4) Defining the minimum requirement for the detection and quantification of genetic disease heterogeneity.
5) Performing real data application in genome-wide genetic data for diseases where subtypes are already known to assess the performances of the method developed.

The selected candidate will be mentored by Dr. Hugues Aschard, but will also work with members of our research group and international collaborators involved in the project. He will have access to all resources at Pasteur, including in particular the High Performance Computing Cluster which includes over 2,000 cores.

QUALIFICATIONS

The position requires advance knowledge in statistics and computer sciences. The applicants should therefore have substantial educational background in Statistics/Biostatistics, Bioinformatics, Computer Science or other relevant disciplines. The call specifically addresses master 2 student and 3rd year engineer school student.

ADDITIONAL INFORMATION

Interested applicants should send their curriculum vitae, a brief cover letter, and contact information from at least one referee (e.g. teacher or previous internship mentor) to Dr. Hugues Aschard (hugues.aschard@pasteur.fr). More information on the Institut Pasteur and the C3BI can be found here https://research.pasteur.fr/en/team/statistical-genetics/ and here https://research.pasteur.fr/en/center/c3bi/.