



INTERNSHIP

Title

Development of a pipeline for multivariate analysis of biological data from antibody discovery and screening

Subject

In the context of antibody discovery and screening, large sets of data including biochemical data, sequences, cellular and functional tests are generated. The specificity of these data is to be of different types: numerical, categorical and text for antibody sequence.

These biological data and the way they are generated will have to be understood by the student to ensure relevant statistical analyses and thus a good interpretation of the results. The work will first focus on literature search to gather multivariate analysis methodologies applicable for these specific sets of data (such as factor analysis and clustering for mixed data, machine learning techniques developed for sequence data, ...) and then explore them on these large datasets. A special attention will be paid to visualization to help communication and sharing of the results. The development of Rshiny app could be evaluated to ensure repeatability over projects and simplified overview of the analyses.

Profile

Master 2 with specialization in biostatistics, bioinformatics or data science
Good knowledge of R and Rshiny
English working skills
Autonomous, organized and able to work in multidisciplinary team

Duration : 6 months

Beginning of the mission : Between February and April 2020

Location : Vitry-sur-Seine (94)

Contact names

The trainee will have co-tutors, one in a laboratory specialized in antibody discovery and screening and the other one in a non-clinical biostatistics team.

To apply, please send resume and cover letter to :

- Valerie MARTIN - Non-Clinical Biostatistics - Vitry-sur-Seine:
valerie.martin3@sanofi.com
- Klervi DESRUMEAUX – Bilogics Reseach – Vitry-sur Seine:
Klervi.Desrumeaux@sanofi.com