2-year postdoctoral position in Statistics and Machine Learning for Biological processes in Lyon, France

Vivian Viallon (International Agency for Research on Cancer, Lyon)
Yohann De Castro (Ecole Centrale de Lyon, Institut Camille Jordan)

Scientific and funding environment

Applications are invited for a 2-year postdoctoral position in statistics and machine learning at the International Agency for Research on Cancer (IARC, World Health Organization, Lyon, France) and Ecole Centrale de Lyon (ECL, Lyon, France). The successful candidate will join the MOBiL project (Multi-omics data integration to investigate biological mechanisms underlying the link between lifestyle behaviors and gastro-intestinal cancers), which is one of the 11 "Projets Structurants" funded within the Shape-Med Lyon initiative (see below for more details).

Under the supervision of Dr. Vivian Viallon (leader of the Biostatistics and Data Integration team at IARC) and Dr. Yohann De Castro (Professor at Institut Camille Jordan, junior member of Institut Universitaire de France), the candidate will develop and study high-dimensional non-linear latent variable models for the integration of multi-omics data. This methodological project is motivated by applications in cancer epidemiology: our models will eventually be applied to the European Prospective Investigation into Cancer and nutrition (EPIC) and UK Biobank studies to *(i)*, characterize human metabolic responses to specific lifestyle exposures (e.g., habitual alcohol intake), while accounting for inter-individual variability of metabolic responses, and *(ii)*, relate these metabolic responses to cancer risk. The epidemiological results will be key to enhance our understanding of biological processes through which lifestyle might influence human health, in particular cancer development.

Proposal and team

Metabolic signatures of lifestyle exposures, such as habitual alcohol intake, have proven useful for the description of metabolic responses to lifestyle exposures and the characterization of biological processes possibly underlying the effects of these exposures on health-related outcomes. Metabolic signatures have usually been constructed using simple statistical models, and ignored inter-individual variability. The main objective of the multi-disciplinary project MOBiL will be to develop more advanced mathematical tools to model inter-individual variability. We will assume that metabolic responses vary across subgroups in the population, characterized by genetic determinants, blood protein levels as well as more standard epidemiological variables, including age and sex. A key challenge is that subgroups to which each subject belongs are unknown. They will be modelled through a latent categorical variable G, which will have to be "predicted" from the data, while estimating the other model parameters.

Models developed in MOBiL will perform multi-omics data integration and dimension reduction through the latent categorical variable *G*. A particular attention will be paid to the trade-off between model complexity to ensure appropriate goodness-of-fit, and tractability to ensure our model can be fitted on real data. To reach this trade-off, which constitutes another key novelty of our approach, different versions of our model will be tested and discussed within the multi-disciplinary team of MOBiL, which combines expertise in statistics and machine learning, but also in biochemistry, molecular epidemiology and cancer epidemiology. The project will also benefit from the expertise the IARC team acquired over the years through the coordination of research projects on the link between metabolic biomarkers with cancer risk, and on the identification of metabolic biomarkers and metabolic signature for specific lifestyle exposures.

Profile of the candidates
Essential or desired selection criteria include:

- A PhD in machine learning, statistics or related fields.
- Strong background within High-Dimensional Statistics, Variational Inference, Bayesian computing, Statistical Learning, or Bio-statistics.
- Strong background within statistical software such as Python or R.
- Proficiency in English (oral and written).
- Strong communication skills.
- Scientific writing skills.

Experience in Biology or Medicine is not required, but experience in biochemistry, molecular epidemiology or cancer epidemiology would be an asset.

Timeframe of the call
The proposal is open and the position can be taken within the next months. Applications for a starting date in 2025 will be also considered. Non-French speakers applicants are welcome, full English job is possible.

Phase 1 (open): Interested candidates should send a full CV, list of publications and motivation letter to viallonv@iarc.who.int and yohann.de-castro@ec-lyon.fr

Phase 2 (invited): Invited candidates will be interviewed online. Up to 2 reference letters can be sent to M. Viallon and M. De Castro before the interview.

Salary and support
Computer, access to large computing facilities, travel expanses and conference fees will be fully covered.

The salary is based on the University Claude Bernard Lyon system, with a monthly net salary ranging from ~1900 euros to ~3100 euros, depending on the level of qualification and experience. Costs related to commuting and mutual insurance can also be covered. Depending on your situation, government tax can be around 3%.

The successful candidate will have the opportunity to teach statistics and machine learning, potentially in English, at Centrale Lyon and/or the EM Lyon Business School.

MOBiL and ShapeMed

The University of Lyon 1, in collaboration with 11 partners, is coordinating the SHAPE-Med@Lyon project (www.shape-med-lyon.fr). The recently awarded MOBiL project, by the SHAPE-Med@Lyon consortium as part of the 2023 seed funding call, aims to develop mathematical models for constructing molecular signatures to study the biological mechanisms through which lifestyle could influence cancer risk. This collaborative project is led by Vivian Viallon from the International Agency for Research on Cancer and Yohann de Castro from the Institut Camille Jordan. The MOBiL project is recruiting a PhD student in statistics to develop latent class statistical models for the integration of multi-omics data and their application to data from large epidemiological cohorts.