2023-06552 - PhD Position F/M [PEPR SN AI4scMED] Single-cell multi-omics data integration for gene regulatory network inference

**Type de contrat :** CDD  
**Niveau de diplôme exigé :** Bac + 5 ou équivalent  
**Fonction :** Doctorant

**A propos du centre ou de la direction fonctionnelle**

The Inria research centre in Lyon is the 9th Inria research centre, formally created in January 2022. It brings together approximately 300 people in 16 research teams and research support services. Its staff are distributed at this stage on 2 campuses: in Villeurbanne La Doua (Centre / INSA Lyon / UCBIL) on the one hand, and Lyon Gerland (ENS de Lyon) on the other. The Lyon centre is active in the fields of software, distributed and high-performance computing, embedded systems, quantum computing and privacy in the digital world, but also in digital health and computational biology.

**Contexte et atouts du poste**

Within the framework of the PEPR project AI4scMED, in which Inria’s team BEAGLE collaborates with research groups in Grenoble, Montpellier, and Paris, you will develop methodological elements and algorithms for the integration of single-cell multi-omics data sets in the context of cell-based precision medicine. Your goal will be to facilitate the subsequent step of gene regulatory network inference, which is crucial for a mechanistic view on the regulatory processes ongoing in healthy and diseased cells.

This is a three-year position for a PhD student. You will be advised by Anton Crombach and Thomas Guyet (HDR) (Inria Centre de Lyon), whilst working together with other members of Work Package 1.2 (N. Varoquaux, A. Cleynen, H. Isambert, L. Cantini) and members of the project AI4scMED. You will be located at Inria’s La Doua site in Villeurbanne / Lyon.

You will benefit from the expertise at BEAGLE regarding bioinformatic data analysis and mathematical and computational modelling of biological systems; and you will profit from world-class knowledge across the many teams involved in the PEPR project AI4scMED. Overall, the project will foster your career by building up cutting-edge expertise on machine learning, high-dimensional data analysis, network inference, and data visualization techniques – areas that are strongly demanded in academia and industry.

**Mission confiée**

Working on Work Package 1.2 of AI4scMED, the recruited person is responsible for the design and implementation of algorithms and user-friendly software that integrate single-cell multi-omics data sets. The goal is to create integrated data sets that can be easily exploited by network inference tools to significantly improve the quality of gene regulatory network models. Specifically, the recruited person will focus on diagonal and mosaic integration, two types of integration where cells and features may be different from one data set to the next. These types of integration have received relatively little attention and only few methods are available for them, each with their advantages and drawbacks. In this context, the recruited person will take initiatives to identify potential improvements in existing methodology and/or to propose novel algorithms and data structures that (1) allow for the application of “classical” analysis tasks such as clustering and visualization, and (2) enable the usage of heterogeneous multilayer networks as developed by L. Cantini and H. Isambert. As a use case, the recruited person will address the 3+1 modal scenario of single-cell transcriptome, chromatin accessibility, and genome conformation combined with spatial omics data (e.g. transcriptomics). The recruited person will have the opportunity to continuously discuss in-depth with AI4scMED experts for each of these data types.

**Principales activités**

The goal of this three-year PhD research position is to develop tools to better integrate heterogeneous single-cell multi-omics data for the study of cellular regulatory processes and their role in disease. This means your main activities are:

- critically assess existing algorithms for single-cell omics data integration, optionally taking into account also spatial transcriptomics data,
- design and implement improved and/or novel algorithms that

---

**Informations générales**

- **Thème/Domaine :** Biologie numérique  
- **Biologie et santé, Sciences de la vie et de la terre (BAP A)**  
- **Ville :** Villeurbanne  
- **Centre Inria :** Centre Inria de Lyon  
- **Date de prise de fonction souhaitée :** 2023-11-01  
- **Durée de contrat :** 3 ans  
- **Date limite pour postuler :** 2023-08-16

---

**Contacts**

- **Équipe Inria :** BEAGLE  
- **Directeur de thèse :** Crombach Antonius / anton.crombach@inria.fr

---

**A propos d'Inria**

Inria est l’institut national de recherche dédié aux sciences et technologies du numérique. Il emploie 2600 personnes. Ses 200 équipes-projets agiles, en général communes avec des partenaires académiques, impliquent plus de 3500 scientifiques pour relever les défis du numérique, souvent à l’interface d’autres disciplines. L’institut fait appel à de nombreux talents dans plus d’une quarantaine de métiers différents. 900 personnes d’appui à la recherche et à l’innovation contribuent à faire émerger et grandir des projets scientifiques ou entrepreneuriaux qui impactent le monde. Inria travaille avec de nombreuses entreprises et a accompagné la création de plus de 180 start-up. L’institut s’efforce ainsi de répondre aux enjeux de la transformation numérique de la science, de la société et de l’économie.

---

**L’essentiel pour réussir**

The applicant should have a strong basis in computer science, mathematics, or biology (MSc degree or equivalent), ideally oriented toward bioinformatic data analysis. An essential quality for this project is a keen interest to apply one’s computational and mathematical skills to help answering biological and biomedical questions. Experience with single-cell omics data analysis and network inference is considered a real asset, as is interest in spatial omics technologies and wanting to
address the challenge of diagonal and mosaic integration in the context of regulatory network inference,

- develop a benchmark suite to test existing and new data integration algorithms,
- convert prototype software into user-friendly software packages, ideally integrating them into the software ecosystem of scverse (Python) and/or seurat (R).

Finally, you disseminate your work through publications in international journals and by presenting it at national and international conferences.

**Compétences**
The project requires skills in computer science, AI/machine learning, mathematics, and biology. A successful candidate has experience in one or more of the following areas: single-cell data analysis, network inference, statistics and machine learning, programming in Python/R. Moreover, affinity with cancer biology or other diseases is considered a real advantage. Good oral and written communication skills in English are essential.

**Avantages**
- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours) + possibility of exceptional leave (sick children, moving home, etc.)
- Possibility of teleworking (after 6 months of employment) and flexible organization of working hours
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Social security coverage under conditions

**Rémunération**
1st and 2nd year: 2082 euros gross salary / month
3rd year: 2190 euros gross salary / month