

2019 Helmholtz – OCPC – Program for the involvement of postdocs in bilateral collaboration projects

PART A

Title of the project: Machine learning methods in single cell genomics

Helmholtz Centre and institute: Helmholtz Zentrum München; Institute of Computational Biology

Project leader: Prof. Dr. Fabian Theis

Web-address: <http://comp.bio>

Description of the project

The Institute of Computational Biology at the Helmholtz Zentrum Munich focuses on the development of models and methods for data analysis in systems biology and systems medicine. One of our research interests is the analysis of heterogeneities in single cell profiles. Single-cell technologies have recently gained popularity in biology due to their ability to resolve heterogeneity in cell populations. This has translational impact e.g. in describing heterogeneities of developmental diseases, cancer or immunological processes. Analyzing high-dimensional single-cell data has its own statistical and computational challenges, and standard tools often cannot be applied. One of our goals is to develop algorithms for non-linear projections of single cell RNAseq data, to for instance decode the regulatory network of early blood development from single-cell gene expression measurements (Moignard et al, Nat. Biotechnol. 2015) and to quantify cellular and molecular properties in large volumes of continuous time-lapse microscopy (Hoppe et al, Nature 2016, Hilsenbeck et al, Nat. Biotechnol. 2016). To identify key driver of cellular decisions in an unbiased fashion, we developed the structure-identification method „diffusion pseudotime“, which allowed us to infer developmental processes from single cell RNA sequencing data (Haghverdi et al, Nat. Meth. 2016).

Within this project we aim at further develop general structure learning algorithms, such as the mentioned diffusion pseudotime, clustering and in particular, topological methods. Eventually, all of these methods aim at generating interpretable maps of cellular datasets, in particular, the Human Cell Atlas (Wolt et al, bioRxiv 2017). Another important aspect the exploitation of deep learning methods for generating well-behaved feature space representations of data, which, for instance, automatically correct for batch and study effects (Eulenberg et al., Nat. Commun. 2017) and the implementation of scalable versions of all of these algorithms (Wolf et al., Genome Biology 2018).

Description of existing or sought Chinese collaboration partner institute

The sought collaboration partner should have

- strong coding knowledge
- basic knowledge of machine learning and statistics
- and ideally some knowledge of biology.

Required qualification of the post-doc:

- PhD in data science, bioinformatics, computer science, statistics, mathematics, physics or equivalent
- strong coding knowledge (preferably Python, C++, some R); basic knowledge of machine learning and statistics, knowledge of biology
- Additional skills: proficient in written and spoken English, being comfortable at interacting with colleagues and experimental partners in an interdisciplinary setting, passion for science and reproducible scientific work