3iA Côte d'Azur
Interdisciplinary Institute for Artificial Intelligence

Ph.D. research topic

- Title of the proposed topic: **Spatial imaging of the human lung from health to pathologies**
- Research axis of the 3IA: **AI for Computational Biology and Bio-inspired AI**
- **Supervisor (name, affiliation, email): Pascal BARBRY, CNRS, barbry@ipmc.cnrs.fr**
- Potential co-supervisor (name, affiliation): **Frédéric Précioso, INRIA**
- The laboratory and/or research group: **IPMC**

Apply by sending an email directly to the supervisor.
The application will include:

- Letter of recommendation of the supervisor indicated above
- Curriculum vitæ.
- Motivation Letter.
- Academic transcripts of a master's degree(s) or equivalent.
- At least, one letter of recommendation.
- Internship report, if possible.

⇨ **All the requested documents must be gathered and concatenated in a single PDF file named in the following format: LAST NAME of the candidate_Last Name of the supervisor_2023.pdf**

- Description of the topic:

Scientific impact

Cellular ecosystems are complex, structured and dynamic. Either *in vivo* or maintained in *in vitro* cultures, the spatial organization of cells contributes to their individual fate. Interactions between neighbouring cells can activate signalling pathways and either drive differentiation mechanisms or trigger pathophysiological mechanisms. Quantifications of gene expression by high-throughput technologies at single cell and spatial resolution have become standards to assess spatialized gene expression. They can better delineate local cellular behaviours that are operating in normal and pathological tissues. The exploration of the fine intercellular relationships that develop between adjacent cells in complex tissues is a central theme of our laboratory. In the context of the Human Cell Atlas Consortium and of the IHU project RespirERA, we are contributing to the creation of comprehensive reference maps of human lung and airway cells in health and diseases [1,2]. These atlases are expected to better understand human health and improve diagnosis, monitoring and treatment of human lung diseases.

Feasibility

The lab is working on a technique called MERFISH [3], an imaging method that is capable of simultaneously measuring the copy number and spatial distribution of hundreds to thousands of RNA species in single cells. A Merscope instrument used for MERFISH is installed in our laboratory since September 2022, i.e. the first instrument in France and the third in Europe. The PhD proposal will tackle several computational challenges to improve biological interpretations of future spatial and single-cell experiments. The student will work on single-cell and spatial transcriptomics datasets

that have been generated by the lab on 5 lung pathologies: chronic obstructive pulmonary disease (COPD), cystic fibrosis (CF), asthma, pulmonary hypertension, and small cell lung cancer (SCLC). The project will develop along the three following axes:

1. **to build an open-access resource of highly specific MERFISH probes against transcripts of interest for each of these 5 pathologies.** Each MERFISH project targets hundreds of genes, through a parallel hybridization of thousands distinct oligonucleotides on tissue sections. Each oligonucleotide is designed to interact specifically with a transcript of interest, after a selection for one of the 3 following reasons: (1) cell-type markers identified in single-cell experiments, (2) differentially expressed genes between different experimental conditions, (3) gene markers for signalling pathways that contribute to important cellular/tissue function(s). Identification of unique regions that do not match with other transcripts is key to ensure the quality of the observation. Specificity and sensitivity of each probe set will be scored, based on criteria previously established by the laboratory for DNA microarray design [4].
2. **to optimize cell boundaries identification in MERFISH experiments.** Resident lung cells display very distinct morphologies (e.g. alveolar type 1 pneumocyte, multiciliated cells, neutrophils, lymphoid cells, etc), which can require specific adaptations to perform their correct segmentation. Segmentation method also needs to be applicable over large tissue areas (1 cm^2). Models derived from state-of-the-art AI approaches (Mesmer [5], Cellpose [6]) will be trained in order to optimize the identification of precise boundary for every cell in the images. A Napari [7] plugin handling the staining and transcriptomics data will adapt the segmentation procedure to each of the 2.500 Fields of Views (FOV) found in a typical MERFISH experiment. On the fly classification of each FOV will be done using optimal parameters of segmentation, which will be further optimized using transcript information. The process will generate polygons representing the different cells and a [cells x genes] matrix. Single cell spatial statistical analysis will then be performed using established Scanpy/Squidpy Python pipelines.
3. **to implement interpretative computational methods.** Datasets will be explored to document the evolution of cell types composition and the expression of target genes. Analyses will be performed at the light of metadata (e.g. position of the biopsy, clinical data, etc), through a Napari window that will be developed using a *Cookiecutter* template. After automatic detection and annotation of spatial region subtypes, profiles of cell types proportion / target genes expression / cell-to-cell interactions will be drawn to highlight trends in those quantifications scores according to remoteness to set identical regions. Differential analysis between pathological and control sections will be used to decipher the determinants of spatial organization for the different cell types in normal and diseased conditions. The use of Graph Neural Networks will be explored to efficiently represent the local structures related to the cell characteristics and the different pathophysiological contexts [8-12]. A last option will explore Transformers, as these super neural models have been recently tested in one work [13], and they outperform most existing models in others multimodal problems.

Originality

All these research directions are extremely new as stated by the date of publications in the references, and they represent a new open field of contributions.

References

[1] Deprez M et al. A Single-Cell Atlas of the Human Healthy Airways. Am J Respir Crit Care Med. 2020 Dec 15;202(12):1636-1645. doi: 10.1164/rccm.201911-2199OC. PMID: 32726565.

[2] Sikkema et al. An integrated cell atlas of the human lung in health and disease. Biorxiv. 2022. doi: https://doi.org/10.1101/2022.03.10.483747

[3] Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. Science. 2015 Apr 24;348(6233):aaa6090. doi: 10.1126/science.aaa6090. Epub 2015 Apr 9. PMID: 25858977; PMCID: PMC4662681.

[4] Le Brigand K et al. An open-access long oligonucleotide microarray resource for analysis of the human and mouse transcriptomes. Nucleic Acids Res. 2006 Jul 19;34(12):e87. doi: 10.1093/nar/gkl485. PMID: 16855282; PMCID: PMC1524919.

[5] Greenwald NF et al. Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. Nat Biotechnol. 2022 Apr;40(4):555-565. doi: 10.1038/s41587-021-01094-0. Epub 2021 Nov 18. PMID: 34795433; PMCID: PMC9010346.

[6] Pachitariu M, Stringer C. Cellpose 2.0: how to train your own model. Nat Methods. 2022 Dec;19(12):1634-1641. doi: 10.1038/s41592-022-01663-4. Epub 2022 Nov 7. PMID: 36344832; PMCID: PMC9718665.

[7] napari contributors (2019). napari: a multi-dimensional image viewer for python. doi:10.5281/zenodo.3555620

[8] Hu, J., Li, X., Coleman, K. et al. SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. Nat Methods 18, 1342–1351 (2021). https://doi.org/10.1038/s41592-021-01255-8

[9] HEYDARI, A. Ali et SINDI, Suzanne S. Deep learning in spatial transcriptomics: Learning from the next next-generation sequencing. bioRxiv, (2022).

[10] Duy Pham et al. stLearn: integrating spatial location, tissue morphology and gene expression to find cell types, cell-cell interactions and spatial trajectories within undissociated tissues. Biorxiv, 2020, doi: https://doi.org/10.1101/2020.05.31.125658

[11] Yuan Y, Bar-Joseph Z. GCNG: graph convolutional networks for inferring gene interaction from spatial transcriptomics data. Genome Biol. 2020 Dec 10;21(1):300. doi: 10.1186/s13059-020-02214-w. PMID: 33303016; PMCID: PMC7726911.

[12] Zhang X, Wang X, Shivashankar GV, Uhler C. Graph-based autoencoder integrates spatial transcriptomics with chromatin images and identifies joint biomarkers for Alzheimer's disease. Nat Commun. 2022 Dec 3;13(1):7480. doi: 10.1038/s41467-022-35233-1. PMID: 36463283; PMCID: PMC9719477.

[13] Zeng Y, Wei Z, Yu W, Yin R, Yuan Y, Li B, Tang Z, Lu Y, Yang Y. Spatial transcriptomics prediction from histology jointly through Transformer and graph neural networks. Brief Bioinform. 2022 Sep 20;23(5):bbac297. doi: 10.1093/bib/bbac297. PMID: 35849101.

Skills: Python, PyTorch, R, Bioconductor