Tools for Spatial Transcriptomics

Submitted by: Marwan Zouinkhi, Stephan Preibisch

Background and current state-of-the-art

Spatially resolved transcriptomics is the description for a range of methods that can identify mRNA readouts for many genes at many locations inside a cell, potentially on a tissue level. The field is currently growing at an immense pace and is transforming biological sciences [1]. Methods for spatial transcriptomics can generally be divided into imaging-based methods such as MERFISH [2] and sequencing-based methods such as SlideSeq [3]. Excitingly, this development leads to a merge of the fields of image and sequence analysis.

We developed two independent software solutions for spatial transcriptomics:

- **Radial Symmetry-FISH (RS-FISH)** [4] is a software that accurately, robustly, and quickly detects single-molecule spots in two and three dimensions, making it applicable to several key assays, including single-molecule FISH (smFISH), spatial transcriptomics, and spatial genomics. It scales from small to very large images using workstations, clusters or the cloud, allows interactive parameter tuning and provides spot visualization of even very large datasets. It is based on ImgLib2, BigDataViewer, the N5 file format and Apache Spark. Currently, it is the only generic tool that can handle very large images. It is already used in the community and at Janelia, for example by the EASI-FISH project, the Sternson lab, or in the sequencing facility.

- **STIM** [5] is an imaging-based computational framework for storing, exploring, visualizing, and processing high-throughput spatial sequencing datasets and can for example be used for robust alignment for slices of SlideSeq and Visium datasets. It is based on ImgLib2, BigDataViewer, and the N5 file format. STIM is not yet used at Janelia, but the community started to adapt it for the handling of sequencing-based spatially resolved datasets, as it is able to transfer image processing concepts such as spatial filtering to irregularly-spaced sequencing datasets.

Janelia is currently a leading institution in image analysis and much of its research is focused on imaging-based methods. However, we anticipate that with the start of 4DCP, spatial transcriptomics will become an integral part of Janelia science as well. Therefore, we think it is important to maintain and further develop tools that can handle and process such data.

Scope of the project

Maintenance, extension, and training for RS-FISH & STIM (3 FTE months)

We propose to maintain, slightly extend, and teach these two frameworks as part of the Open Science Software Initiative. This includes:

- code-refactoring
- additional documentation

- **STIM**:
  - consolidation of the API
  - support of new file formats for reading (Visium) and writing (HDF5)
- **RS-FISH**:
  - better support of MERFISH data
  - generalization to other spot detection methods that can be implemented for small blocks of data (e.g. Srini's DL-based smFISH proposal) since a lot of the code is about data handling, visualization and parallelization
  - time permitting RS-FISH should be combined with state-of-the-art registration frameworks like BigStitcher [6] to support multiplexed, multi-round smFISH datasets

Documentation & outreach

We plan to teach RS-FISH and STIM at conferences, workshops and within Janelia. We will extend the documentation and plan to release an almost finished bioRxiv paper for STIM.

Required skillset

This project will require a part-time developer (e.g. an SCA or an engineer) that will be supervised and supported by Stephan Preibisch.

Estimated impact

RS-FISH is very new and already used by the community (determined by twitter feedback, requests, and bioRxiv "*it's in the top 5% of all research outputs ever tracked by Altmetric*") because of its unique capabilities compared to other existing spot detection software (large datasets, speed, robust, interactive parameter tuning). We expect that dedicated teaching and maintenance will further improve usage of the software.

STIM is currently not widely used, but we are convinced that releasing the bioRxiv together with teaching will boost its use. STIM has the unique feature that it efficiently combines the worlds of image and sequence analysis, thus providing valuable tools to the sequencing community.

References

[1] Marx, V., "Method of the Year: spatially resolved transcriptomics", *Nat Methods* **18,** 9–14 (2021).

[2] K.H. Chen, A.N. Boettiger, J.R. Moffitt, S. Wang, X. Zhuang, "Spatially resolved, highly multiplexed RNA profiling in single cells", *Science* **348**, aaa6090 (2015).

[3] Rodriques, S. et al., "Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution", *Science* **363**, aaw1219 (2019).

[4] Ella Bahry, Laura Breimann, Marwan Zouinkhi, Leo Epstein, Klim Kolyvanov, Xi Long, Kyle I S Harrington, Timothée Lionnet, Stephan Preibisch, "RS-FISH: Precise, interactive, fast, and scalable FISH spot detection", *bioRxiv* 2021.03.09.434205 (2021).

[5] https://github.com/PreibischLab/STIM

[6] Hörl D, Rojas Rusak F, Preusser F, Tillberg P, Randel N, Chhetri RK, Cardona A, Keller PJ, Harz H, Leonhardt H, Treier M, Preibisch S., "BigStitcher: reconstructing high-resolution image datasets of cleared and expanded samples", *Nature Methods* **16**, 870–874 (2019).