

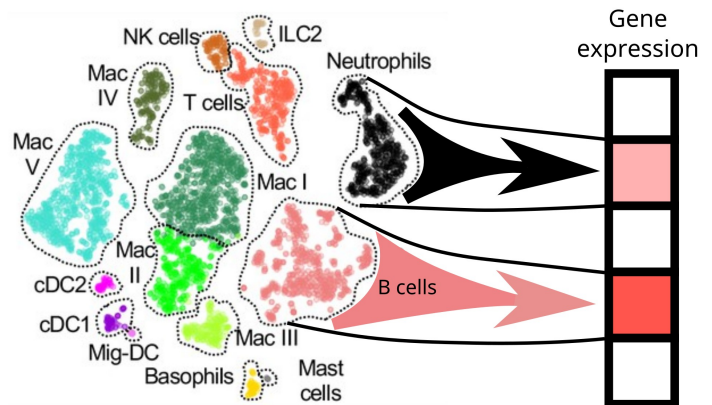


**UNSW**  
SYDNEY

Fabio Zanini  
Low Cancer Research Centre, Level 2, Room 211  
University of New South Wales  
NSW 2052  
Australia

**Title:** Light and scalable cell atlas approximations for machine learning

**Project description:** Cell Atlases are massive and extremely granular data sets describing virtually all cells of an entire organism at once (e.g. human), usually in terms of gene expression. While data on several organisms are theoretically publicly available online, in practice their exploration is challenging because they are both big in size and inconsistent in format. Moreover, even when properly organised, a full cell atlas contains an overwhelming amount of information such that typical data analyses approaches only scratch the surface and fail to deliver useful insight.



This project aims at constructing (1) approximations of cell atlases based on a lossy compression algorithm and (2) interfaces that enable both machines and human to browse these approximations rapidly and effectively. Data and backend software engineering will be used to optimise the compression and design the API, while frontend engineering and natural language processing (NLP) will be used to design the human interface to be used by biologists and medical doctors.

**Requirements:** Background in computational biology, data engineering, data science, software engineering and/or machine learning. A proven record of contribution to open source projects including the ability to use git and GitHub proficiently is expected. The ideal applicant should be motivated, fast-learning, and passionate about science. Basic knowledge of biology or medicine is a plus.

**Contract timeline:** 1 year plus 1 additional year conditional on performance review.

**Location:** Sydney, Australia. Remote work considered for exceptional candidates.

**Salary:** Approximately Level A Step 5 equivalent (~\$100,000/year + benefits).

**Contact:** Fabio Zanini (<http://fabilab.org>)