

Title: Evolutionary machine learning with compressed cell atlases

Supervisors: Fabio Zanini (http://fabilab.org), joint supervisor TBD

Project timeline: 3.5 years.

Project description: A "cell atlas" is a vast data set that provides an extremely detailed view of the cellular diversity ²⁰ of an entire organism. During the last few years, scientists ¹⁵ have collected data from tens of organs covering not only ¹⁰ humans but also several other species including apes, ⁵ mouse, fish, jellyfish, and plants. Cell atlases can revolutionise biomedicine by providing instantaneous ⁰ knowledge about the types of cells found in a tissue, their ⁻⁵ abundance, and their function. However, cell atlases are-10 currently difficult to use because their analysis and manipulation requires advanced bioinformatic skills, a



powerful workstation, and a deep understanding of organismal biology.

This project aims at lifting the barriers for the analysis of cell atlases. The student will develop a completely new approach to cell atlas analysis that relies on lossy data compression, machine learning, and evolutionary biology. For instance, the student will build twodimensional embeddings of single cells aided by a cell atlas (see figure), and develop algorithms that combine graph clustering and embedding into a single, coherent step to enable intuitive browsing of cell atlas information by both human and machines.

Skills you will learn: The student will learn to perform world-class research in data science, software engineering, and computational biomedicine. Graph algorithms, data engineering, and machine learning will be mixed organically on a daily basis. The student will be guided in all aspects with a clear route towards scientific independence.

Requirements: Background in computational biology, data science, and/or computer science. The student should be motivated, fast-learning, and passionate about science. Knowledge of cell biology is a plus.

Contacts: For inquiries write to fabio.zanini@unsw.edu.au