

## Leveraging public single cell data for bench biologists

---

**Principal supervisor** Wendi Bacon (Life, Health & Chemical Sciences)

**Co-supervisor** Paul Mulholland (Knowledge Media Institute)

**Location** The Open University, Milton Keynes, United Kingdom

**Full-time only**

**Duration & Funding** 3 year 3 month studentship as part of EPSRC Doctoral Training Partnership; Stipend £18,622 per annum; Training grant £4,500

**Application due date:** Jan 25, 2024

**Notification of shortlisting:** Feb 8, 2024

**Interview: Feb 22, 2024 on Microsoft Teams (can be flexible on date if needed)**

**Final Funding Decision:** Late April/early May 2024. This is part of a pooled EPSRC-DTP process, so the selected applicant will be put forward to a reviewing panel in April for final decisions. Applicants will be notified if they are selected, and will be informed of the panel decision afterwards.

**Start date:** October, 2024

**Science-related enquiries:** [wendi.bacon@open.ac.uk](mailto:wendi.bacon@open.ac.uk)

**Process-related enquiries:** [STEM-lhcs-phd@open.ac.uk](mailto:STEM-lhcs-phd@open.ac.uk)

---

**Research area/keywords:** Human computer interaction, Single cell RNA-sequencing; Bioinformatics; Galaxy Project; Computer science

## Project background and description

**Leveraging public single cell data:** Single cell technologies allow scientists to identify molecules within a single cell across thousands of cells. Such technologies have flourished - the Human Cell Atlas currently boasts data from nearly 50 million cells while the Single Cell Expression Atlas contains 10 million cells across multiple species. Such atlases contain vast quantities of data available for researchers, if they can use it.

The most common method for leveraging public data is to simply replicate an analysis but identify different points of interest. What if bench biologists could do more than explore or replicate public data? What if they could create something entirely new?

Such analyses exist - RNA deconvolution uses single cell reference datasets to infer cell type proportions from bulk RNA-seq samples. Spatial deconvolution infers cellular location within a tissue. CellPhoneDB searches data for receptor-ligand network analysis to identify how cells interact. Drug2cell allows scientists to find cells which are targeted by drugs, or find drugs which target cells of interest.

**Barriers to usage:** For bench biologists, analyses beyond replication are rarely achievable. While tutorials exist to guide users through a basic analysis of single cell data – which itself can be replete with barriers such as computing environment differences or datatype heterogeneity – more advanced inference skills that combine datatypes are rarely reproducible by a scientist without programming skills.

**Galaxy:** The Galaxy project is an open source, web-based platform for data intensive biomedical research that enables scientists sans programming skills to perform big data analysis. Launched in 2005, it has been cited in over 12,000 publications, contains nearly 10,000 bioinformatics tools, and over 400,000 users across the globe in academia and industry. Tool wrapping is laborious, however, and is often a limiting factor for use in cutting-edge research.

**Project Overview:** In this project, the PhD student will combine the fields of human-computer interaction, single cell bioinformatics, and public data inference to build user-friendly Galaxy workflows that allow bench biologists to explore their data in a context of petabytes of public references. The student will: review available tools that combine or repurpose public data in unique ways; interview bench biologists and bioinformaticians to identify user needs; investigate AI tools to automate tool development; wrap necessary tools and workflows into the Galaxy interface; create relevant documentation, including training materials; and test their analyses with users as case studies to ensure end-to-end capability for a bench biologist performing a study. Embedded within this user design strategy for enabling analysis will be an emphasis on intuiting for the user the procedural knowledge of the analysis, i.e. how the software works, to ensure that biologists are both able to understand what the software does to get them their answer, but also the steps required to perform the analysis.

## References

1. Galaxy Community. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Res.* **50**, W345–W351 (2022).
  2. Moreno, P. *et al.* User-friendly, scalable tools and workflows for single-cell RNA-seq analysis. *Nat. Methods* **18**, 327–328 (2021).
  3. Wang, X., Park, J., Susztak, K., Zhang, N. R. & Li, M. Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. *Nat. Commun.* **10**, 380 (2019).
  4. Garcia-Alonso, L. *et al.* Single-cell roadmap of human gonadal development. *Nature* **607**, 540–547 (2022).
  5. Kanemaru, K. *et al.* Spatially resolved multiomics of human cardiac niches. *Nature* **619**, 801–810 (2023).
-

## Eligibility

1. Applicants will ideally have a First Class or Upper Second undergraduate degree or Masters degree (or equivalent experience) in an area related to the project (e.g. computer science; bioinformatics)
2. The student would be required to live in the UK and within commuting distance to The Open University in Milton Keynes.
3. This funding is for UK students only (see the [definition of UKRI home student](#))
4. Programming languages, especially Python

## Desirable Criteria

1. Independent thinking skills & initiative
2. Scientific writing
3. Working well in a team, especially remote working
4. Experience working with Open Source projects

We are committed to widening participation and awarding PhD studentships to a diverse community of applicants. We particularly welcome applications from under-represented groups. Equal Opportunity is University policy.

---

## How to apply

Please check this page for application entry requirements:

<https://www.open.ac.uk/postgraduate/research-degrees/degrees-we-offer/doctor-of-philosophy-phd>

Please submit to [STEM-LHCS-PHD@open.ac.uk](mailto:STEM-LHCS-PHD@open.ac.uk) an:

- application form, and
- 2-page (A4) personal statement outlining your suitability for the studentship, what you hope to achieve from the PhD and your research experience to date

You do not need to submit a research proposal.

Information and the application form is found here:

<https://www.open.ac.uk/postgraduate/research-degrees/how-to-apply/mphil-and-phd-application-process>. Note that as part of the application form, you will be asked to submit further documents (CV, degree transcripts, etc.)

---