ICR The Institute of Cancer Research

PhD Project Proposal

Funder details

Studentship funded by:	Breast Cancer Now

Project details

Project title:	Computational analysis of spatial dynamics in breast cancer

Supervisory team

Primary Supervisor:	Esther Arwert
Associate Supervisor(s):	Syed Haider, Luis Zapata
Secondary Supervisor:	Rachael Natrajan

Divisional affiliation

Primary Division:	DBC
Primary Team:	Functional Tumour Immunology Embedded within Syed Haider's Data Science Team
Site:	Chelsea

Project background

Breast cancer remains a leading cause of cancer-related deaths globally, with triple-negative breast cancer (TNBC) presenting particular therapeutic challenges due to its aggressive nature and limited treatment options. While recent advances in immunotherapy have shown promise, response rates vary significantly among patients. Understanding the mechanisms behind treatment response requires detailed analysis of the complex interactions within the tumour microenvironment (TME).

Working within the Data Science team, the student will develop computational skills required for the basic analysis and complex integration of multi-modal datasets.

Dr. Arwert's laboratory investigates how cancer-associated fibroblasts (CAFs) influence immune cell function, particularly T cells, in the breast cancer TME. One of their experimental approaches is the use of repeat sampling techniques - Fine Needle Aspiration (FNA) and Core Needle Biopsies (CNB) - which allow temporal tracking of immune tumour evolution in individual tumours over time. This clinically relevant sampling approach generates sequential datasets from the same tumour, providing essential information about how the TME changes over time. Their work generates extensive multi-modal datasets using contemporary technologies including single-cell RNA sequencing (scRNA-seq), spatial transcriptomics, and multiplex immunohistochemistry. Whilst each of these methods provides valuable information about cellular composition and spatial organisation within tumours, the integration of these different data types, particularly across timepoints, is essential to build a comprehensive understanding of how the TME evolves during cancer development and treatment.

This PhD project will be supervised jointly by Dr. Arwert and Dr. Haider, combining expertise in experimental cancer biology with computational methods. Dr. Haider leads the Breast Cancer Research Data Science team, which specialises in developing computational methods for analysing complex cancer datasets. The team has established expertise in single-cell genomics, spatial transcriptomics, and statistical machine learning applications in cancer research. Their computational work has contributed to several key breast cancer studies, including the identification of treatment resistance breast cancers and biomarkers associated with these.

Project aims

- Develop computational approaches to integrate temporal single-cell RNA sequencing data from repeat biopsies with spatial transcriptomics to map the evolving tumour microenvironment.
- Create computational approaches to characterise and quantify CAF-T cell interactions across treatment timepoints.
- Build and benchmark predictive models of treatment response based on early TME changes using statistical machine learning.
- Design interactive visualisation tools for exploring longitudinal multi-modal spatial and single-cell data.

Research proposal

This is a computational-only project.

This project will develop computational approaches for integrating single-cell and spatial transcriptomics data from the breast cancer tumour microenvironment (TME), with future extension to temporal analysis as new datasets become available. The primary aim is to map how cancer-associated fibroblasts (CAFs) and T cells interact within the spatial context of breast tumours, and how they spatially interact with cancer cells.

1. Initial Data Integration and Method Development
The first phase will focus on integrating existing single-cell RNA sequencing data from public breast cancer datasets with newly generated spatial transcriptomics data from the Arwert laboratory. This will establish foundational methods for:

- Processing and quality control of spatial transcriptomics data
- Integration of single-cell and spatial data modalities
- Mapping cell type specific signatures from scRNAseq onto spatial coordinates
- Validation using control datasets with both spatial and single-cell measurements

We will compare established methods (e.g. Seurat) while developing bespoke extensions to these approaches for cross-platform integration. The spatial transcriptomics data will provide information about the physical organisation of the TME, while single-cell data will offer high-resolution cellular phenotyping. These modalities will be integrated along with matched pathology images/annotations.

2. Spatial quantifying of CAF-T Cell Interactions
Using the integrated datasets, we will:

- Develop methods to identify significant spatial associations between cell types
- Create methods to define cellular neighbourhoods and interaction patterns thereof
- Implement statistical frameworks to test interaction strength (e.g. Delaunay triangulation)
- Build approaches to validate computational predictions using multiplex immunohistochemistry data
- Benchmark against existing breast cancer spatial transcriptomics datasets

- Compare findings with experimental co-culture results

A key innovation will be developing metrics that capture spatial relationships between different cell types, particularly focusing on CAF-T cell interactions. This will allow us to identify distinct interaction patterns and correlate these with tumour features and treatment response.

3. Extension to Temporal Analysis
As repeat biopsy data becomes available from FNA and CNB samples, we will extend our methods to incorporate temporal dynamics. This will involve:

- Implementation of quality metrics specific to repeat sampling
- Adapting existing methodology for temporal datasets
- Creating methods to track changes in spatial relationships over time

4. Implementation of generic publication quality software
To ensure reproducibility and useability, we will:

- Create standardised workflows for data processing
- Implement quality control metrics at each analysis stage
- Develop visualisation tools for exploring results
- Release code with thorough documentation

Expected Outcomes:

Technical Deliverables:

- Expertise in using HPC and bash
- Validated pipelines for single-cell and spatial data integration with built-in signatures of CAF and T-cells
- Methods for quantifying cell-cell interactions in spatial context
- Framework for spatio-temporal analysis

Biological Insights:

- Spatial maps of CAF-T cell interaction patterns
- Identification of key spatial organisations in the TME
- Understanding of CAF-T cell relationships in different tumour regions
- Testable hypotheses about CAF influence on T cell function

This is an interdisciplinary project and work will be conducted within the Breast Cancer Research Data Science team, in close collaboration with experimental laboratories. This environment ensures computational rigour while maintaining biological relevance through continuous experimental validation.

The methods developed will advance our understanding of breast cancer biology while providing tools applicable to other systems where spatio-temporal analysis is crucial. Through this work, we aim to better understand the spatial organisation of the TME and how this affects tumour progression and treatment response.

Literature references

Candidate profile

Note: the ICR's standard minimum entry requirement is a relevant undergraduate Honours degree (First or 2:1).

Pre-requisite qualifications of applicants:	• Candidates must have a First or 2:1 Honours degree or a master's in computational biology, biology, or a related discipline or equivalent in a quantitative subject. • Experience in at least one analytical programming language
Intended learning outcomes:	• Proficiency in using HPC, bash and version control system (git) • Advanced programming skills in R and Python for biological data analysis • Expertise in processing and analysing single-cell RNA sequencing data • Proficiency in spatial transcriptomics analysis • Experience in developing and implementing statistical methods for spatial analysis • Experience interdisciplinary collaboration between wet-lab and computational biology • Develop critical thinking skills in cancer immunology and tumour evolution research • Develop scientific presentation and writing skills

Advertising details

Project suitable for a student with a background in:	☒ Biological Sciences ☒ Physics or Engineering ☐ Chemistry ☒ Maths, Statistics or Epidemiology ☒ Computer Science