Mosaic single cell data integration

Dr Shila Ghazanfar
Lecturer, ARC DECRA
School of Mathematics and Statistics, Faculty of Science

BIRS 2023 Single-Cell Plus – Data Science Challenges in Single-Cell Research

THE UNIVERSITY OF SYDNEY

CRICOS 00026A

Mission: to use technical capacity and methodological creativity to solve emerging data problems in biomedical research

Sydney Precision Data Science Centre

sydney.edu.au/science/data-science

Australian Government

Australian Research Council

Chan Zuckerberg Initiative

Why perform single cell data integration?

- Joint visualization
- Joint supervised learning
- Joint unsupervised learning of common clusters
- Cell abundance hypothesis testing
- Imputation of missing modalities
- Joint bespoke analysis (e.g. pseudotime inference)

Horizontal and vertical single cell data integration

nature biotechnology

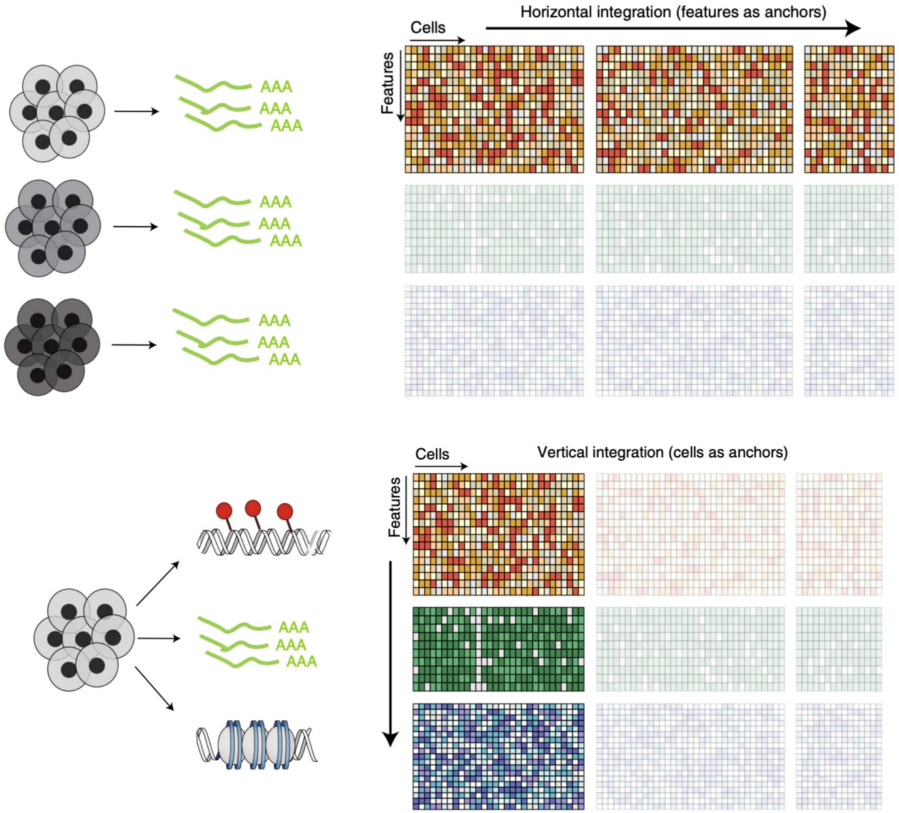
REVIEW ARTICLE

https://doi.org/10.1038/s41587-021-00895-7

Check for updates

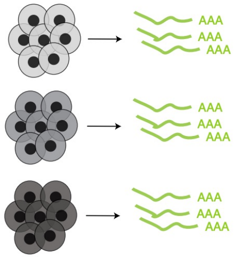
Computational principles and challenges in single-cell data integration

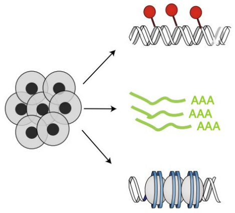
Ricard Argelaguet[1,2], Anna S. E. Cuomo[1,3], Oliver Stegle[3,4,5] and John C. Marioni[1,3,6]

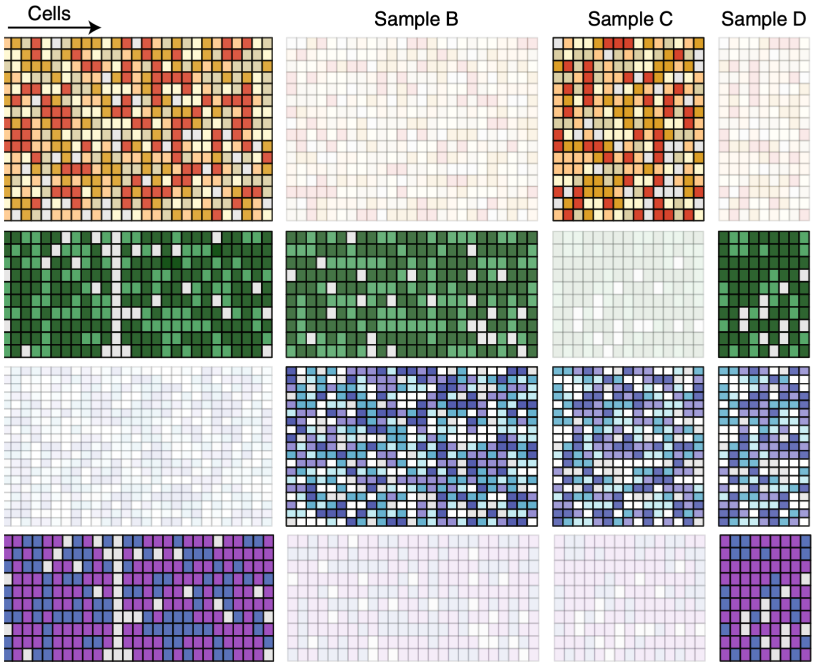


Horizontal integration (features as anchors)
Cells
Features
AAA
AAA
AAA
Vertical integration (cells as anchors)
Cells
Features
AAA
AAA
AAA

Mosaic single cell data integration

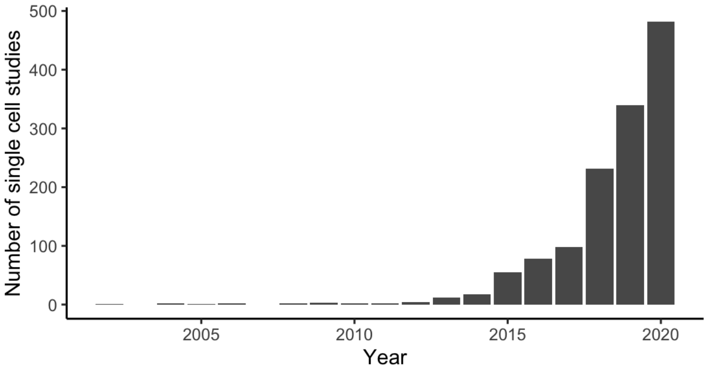


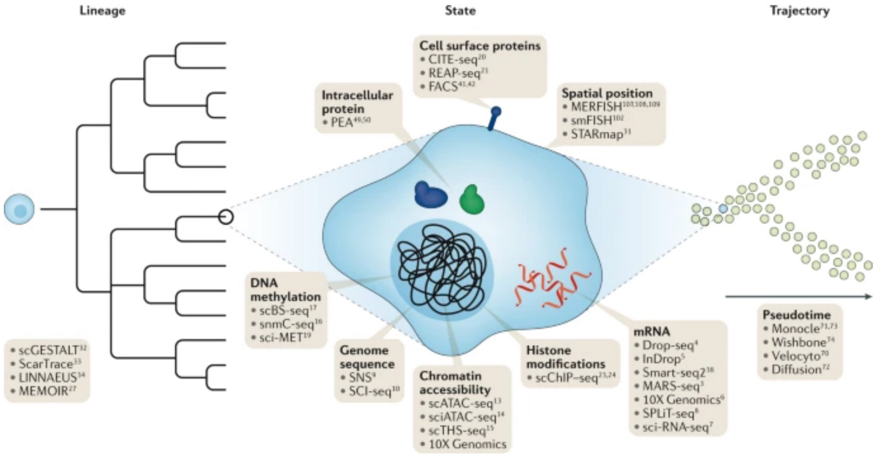




Why do we need to consider mosaic data integration?

Increase in the ***number*** and ***variety*** of single cell technologies





Stuart & Satija (2019)

Svensson et al (2020)

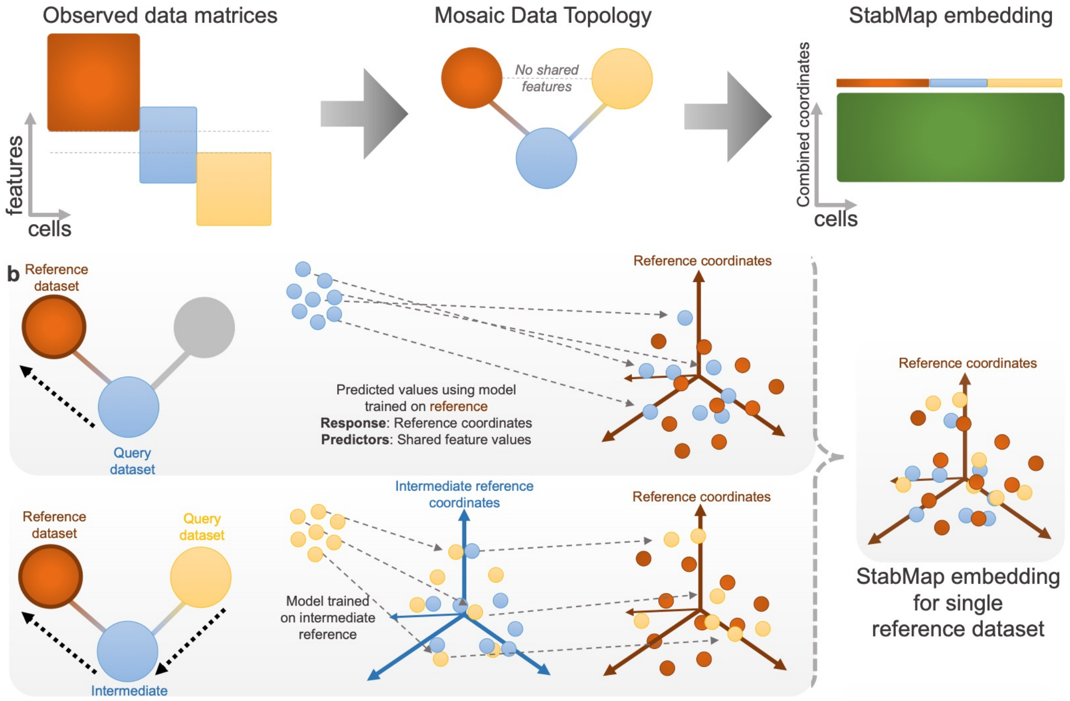
Goals for mosaic single cell integration

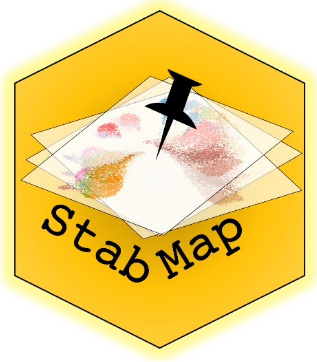
- Develop a technique that performs mosaic data integration, using information derived from non-intersecting features.

- Enable indirect mosaic data integration, by first extracting shared feature relationships among datasets.

- Incorporate prior information from cell labels in the mosaic data integration.

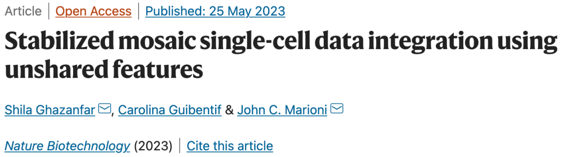
StabMap: Stabilised mosaic single cell data integration using unshared features





John Marioni

Carolina Guibentif



Article | Open Access | Published: 25 May 2023

Stabilized mosaic single-cell data integration using unshared features

Shila Ghazanfar, Carolina Guibentif & John C. Marioni

Nature Biotechnology (2023) | Cite this article

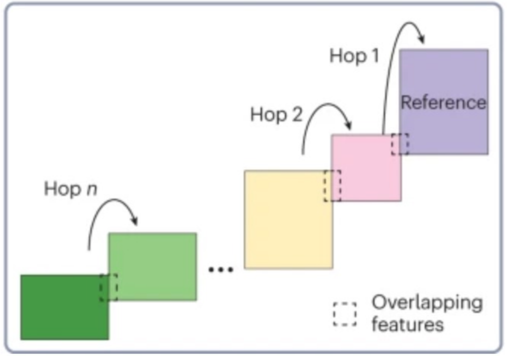
https://github.com/MarioniLab/StabMap

StabMap: requirements, features, and underlying assumptions

Requirements:
- mosaic data topology must be **connected**

Features:
- General, implemented for any connected topology
- User weighting of reference coordinates contribution
- Deterministic and linear
- Requires normalization of input data
- Can be paired with other horizontal and vertical integration

Assumptions:
- No confounding of biological signal between datasets
- Enough biological information captured among shared features

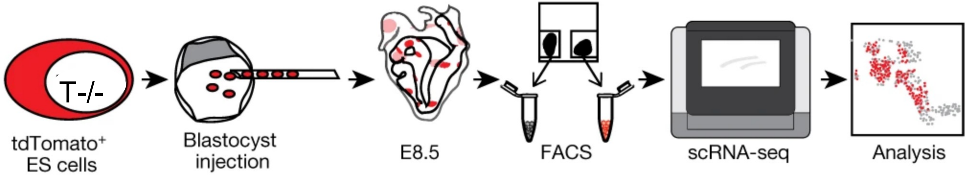


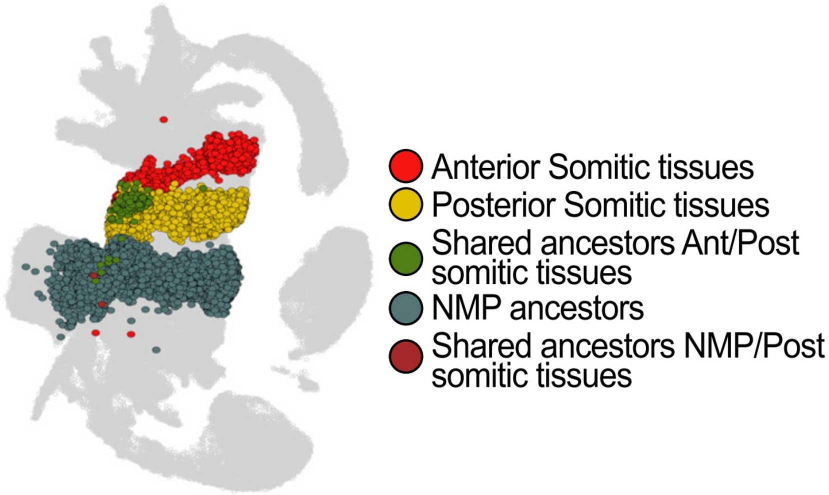
From Li et al, Nature Biotechnology

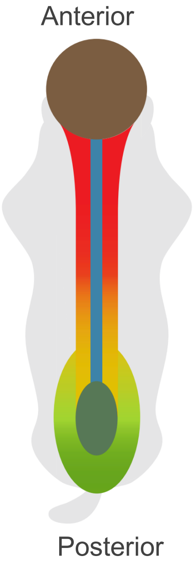
Case study:

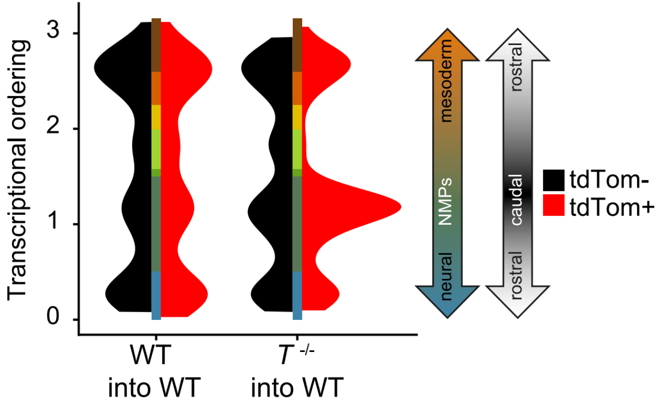
Mapping mutant cells to spatial omics reference

scRNA-seq profiling of Brachyury mutant chimera





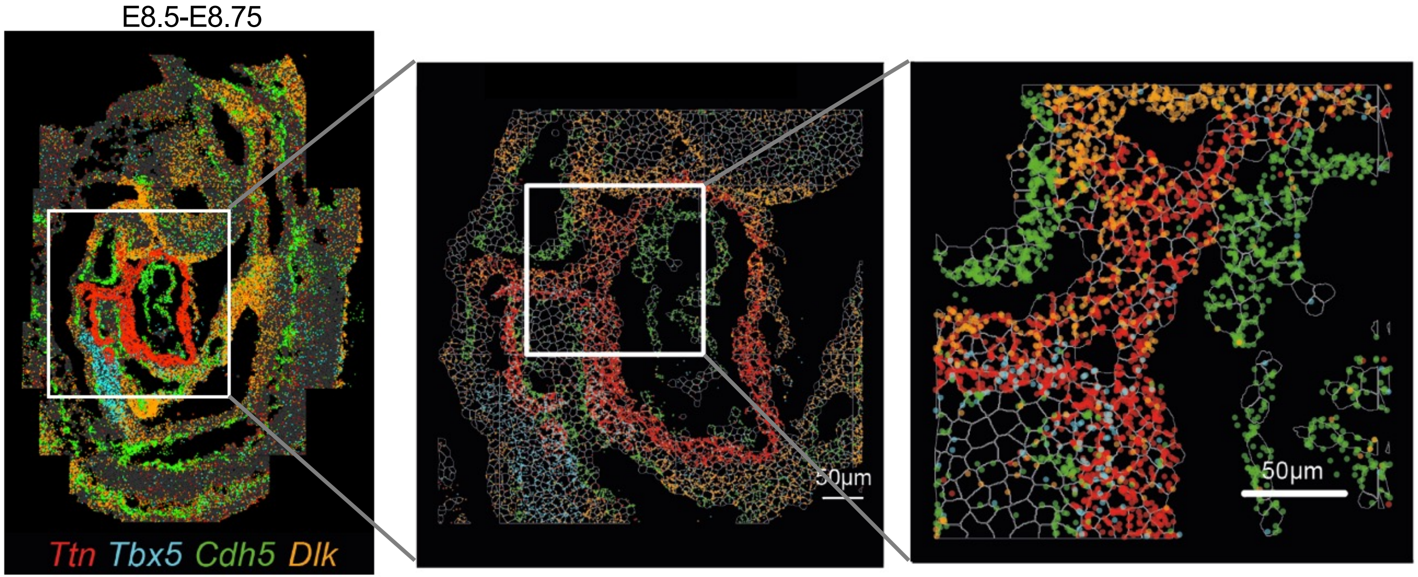




Atlas from Pijuan-Sala et al, Nature (2019)

Guibentif et al, Developmental Cell (2021)

Molecule-resolved early organogenesis spatial mouse atlas

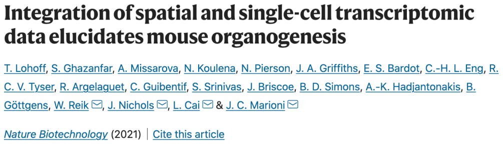


Whole organism section *Tissue* *Subcellular*

Tim Lohoff
Babraham Inst

Long Cai
Caltech

John Marioni
Cambridge

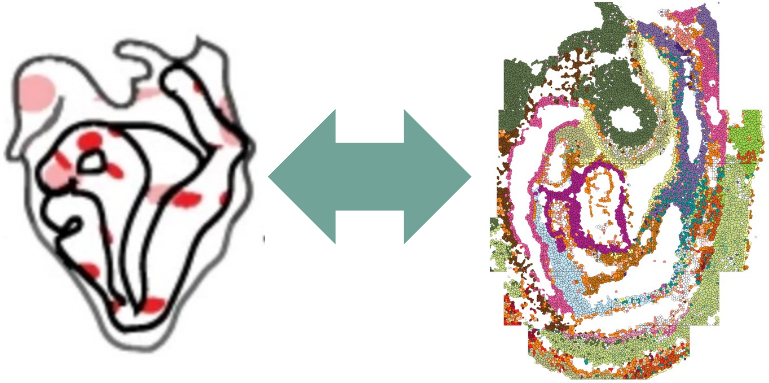


Integration of spatial and single-cell transcriptomic data elucidates mouse organogenesis

T. Lohoff, S. Ghazanfar, A. Missarova, N. Koulena, N. Pierson, J. A. Griffiths, E. S. Bardot, C.-H. L. Eng, R. C. V. Tyser, R. Argelaguet, C. Guibentif, S. Srinivas, J. Briscoe, B. D. Simons, A.-K. Hadjantonakis, B. Göttgens, W. Reik ✉, J. Nichols ✉, L. Cai ✉ & J. C. Marioni ✉

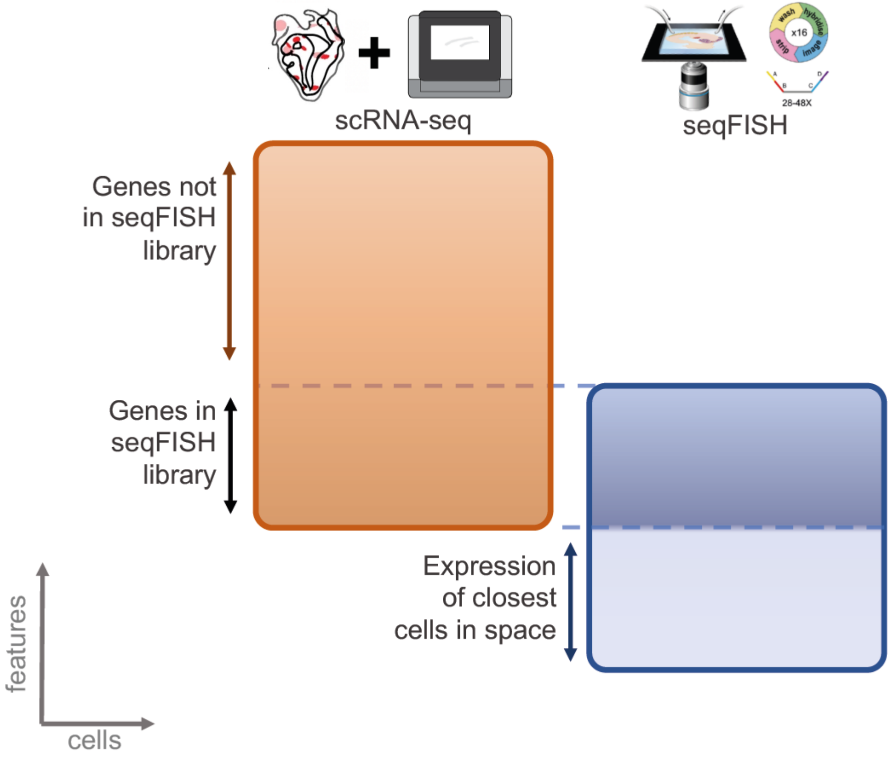
Nature Biotechnology (2021) | Cite this article

Mosaic integration problem



Brachyury (T) chimera

Wild type



scRNA-seq

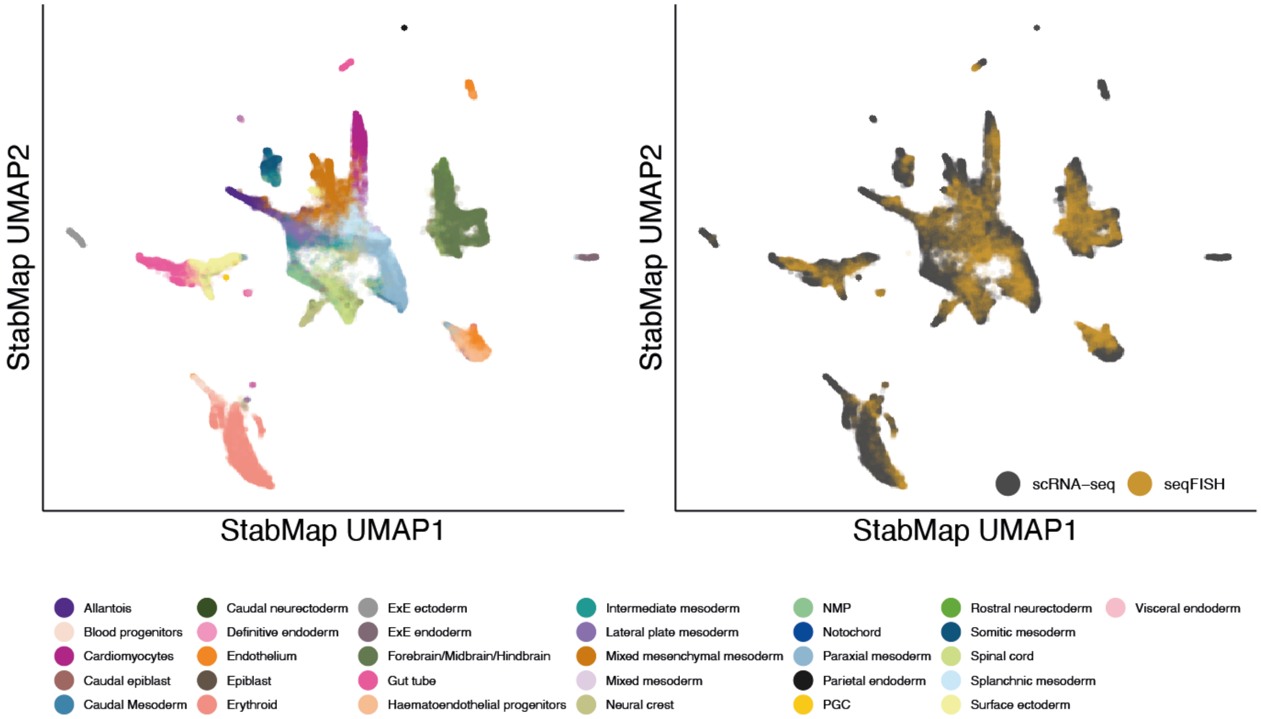
seqFISH

Genes not in seqFISH library

Genes in seqFISH library

Expression of closest cells in space

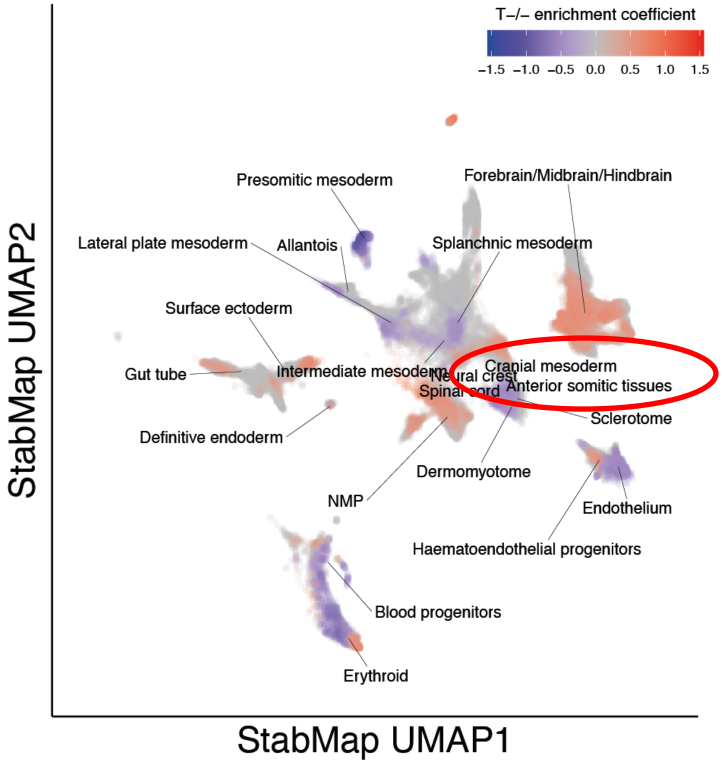
Mosaic integration of scRNA-seq and seqFISH cells



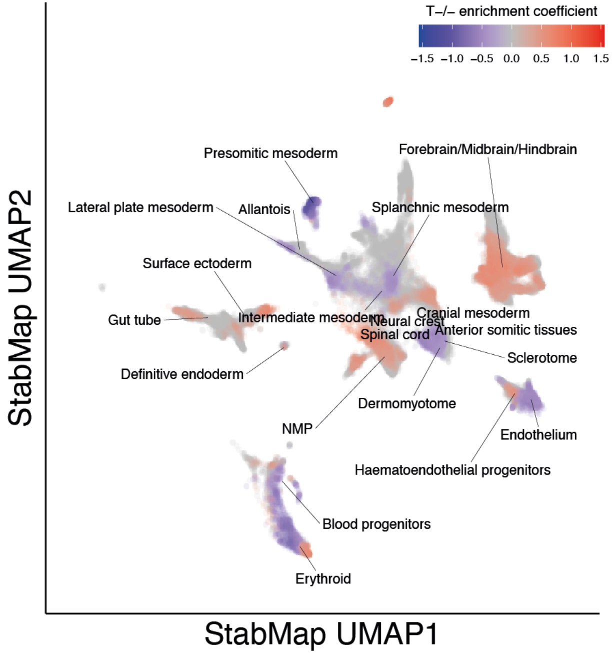
Test for overabundance of mutant cells near seqFISH-resolved cells

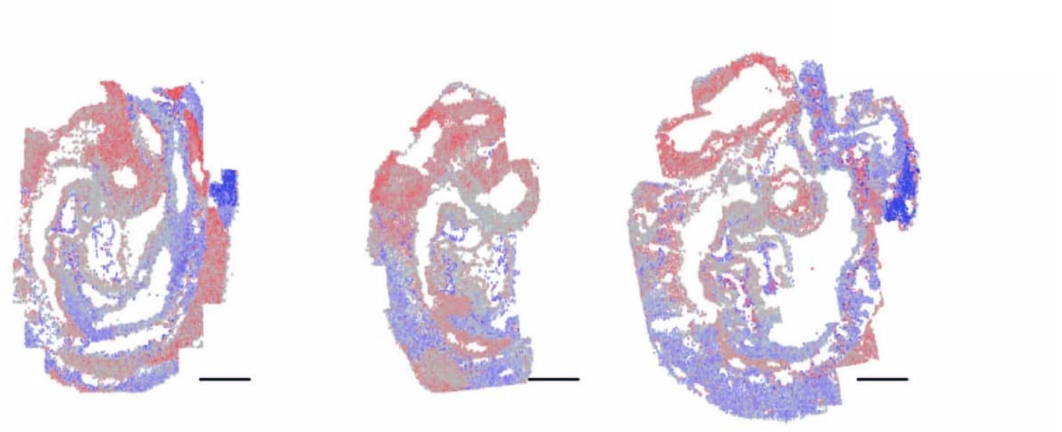
Testing approach: For each seqFISH-resolved cell:

- identify the nearest K (=1000) scRNA-seq resolved cells from within the StabMap embedding;
- Calculate proportion of WT/T- cells among the K nearest
- Compare to global proportion of WT/T- via binomial test;
- Report T- enrichment coefficient.

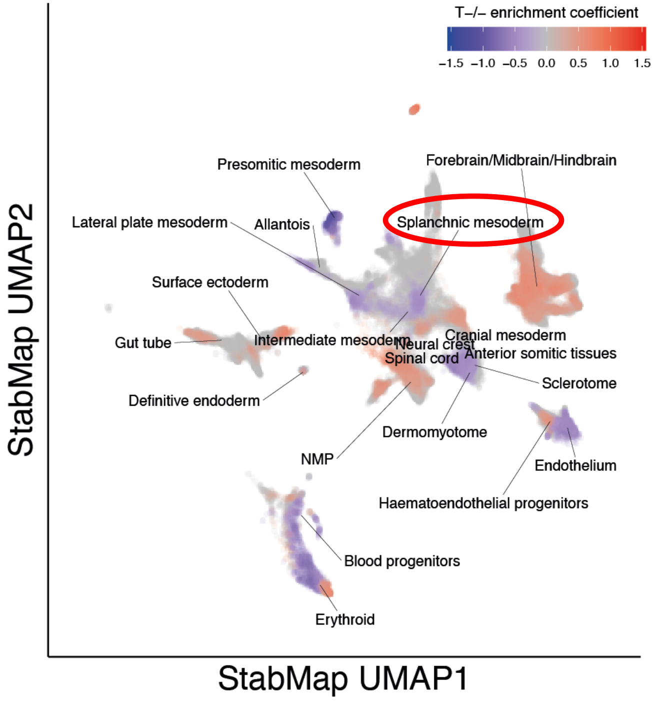


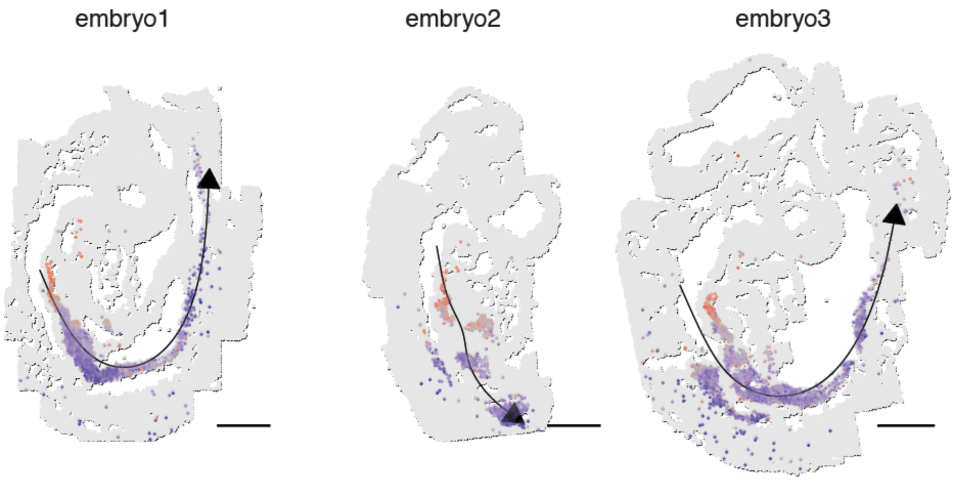
Test for overabundance of mutant cells near seqFISH-resolved cells





Anterior enrichment implicated in other mesoderm type





Mosaic single cell data integration

Data Science Challenges

Data Science challenges: single cell mosaic data integration

- **Relevant and efficient extraction of features**
- How to best combine with vertical and horizontal integration
- **Estimating errors across multiple ‘hops’**
- Challenge of ‘diagonal’ integration
- Potential to go beyond underlying linear model

Thank you!

Carolina Guibentif
University of Gothenburg

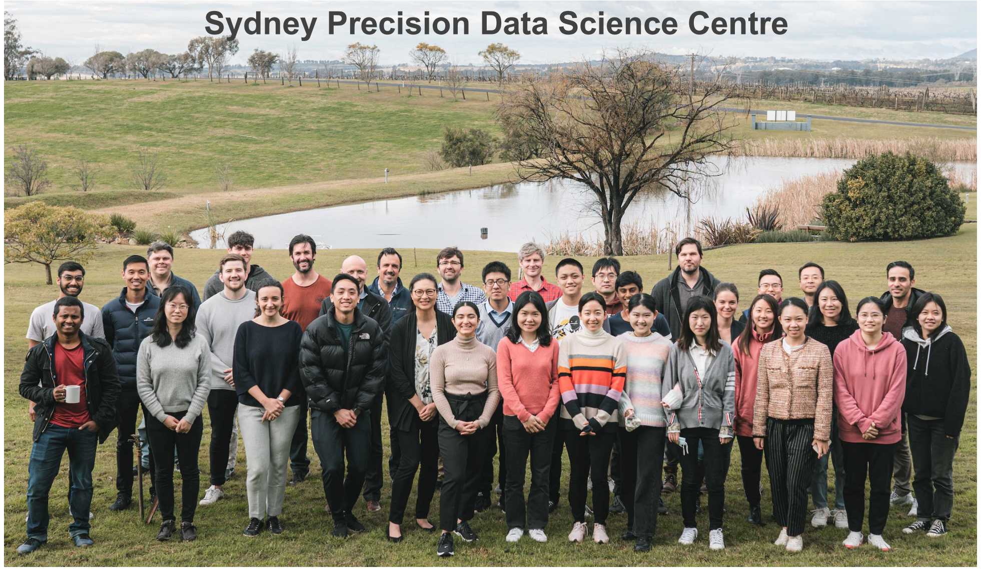
John Marioni
Genentech, CRUK-CI,
EMBL-EBI

Mike Morgan
University of Aberdeen

Karsten Bach
ETH Zurich

All members of the Marioni Lab

Sydney Precision Data Science Centre



sydney.edu.au/science/data-science

Australian Government
Australian Research Council

Chan Zuckerberg Initiative

THE UNIVERSITY OF SYDNEY