Integrative analysis of breast cancer survival based on spatial features

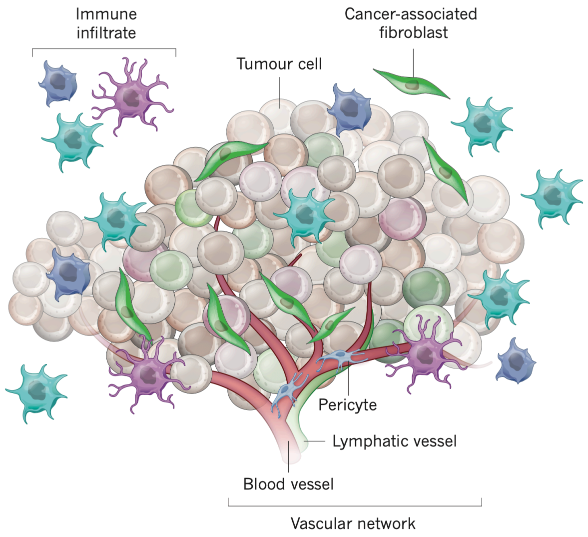
Presented by
Yingxin Lin

School of Mathematics and Statistics
The University of Sydney

THE UNIVERSITY OF SYDNEY

Motivation

Tumour microenvironment



Clinical features

Grade

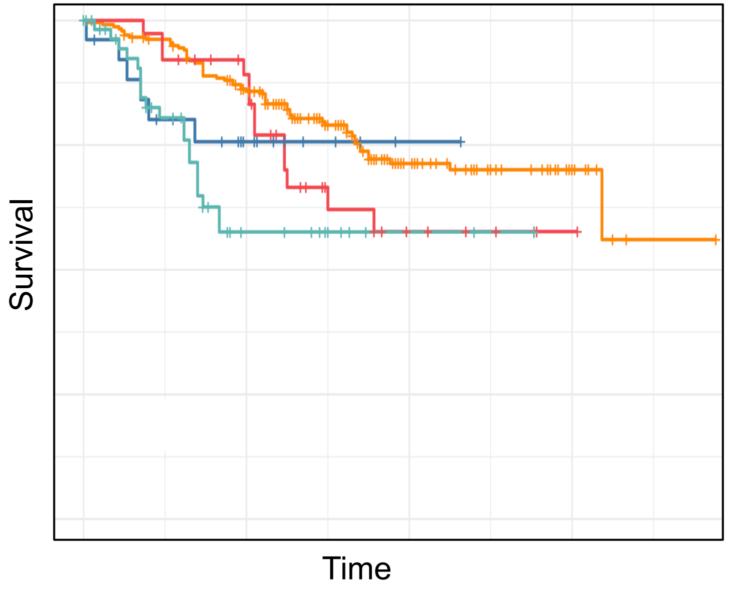
Stage

Tumour infiltration score

Clinical type

Age

Patient survival



Non-spatial features

Cell type composition

Marker

Cell state

Spatial features

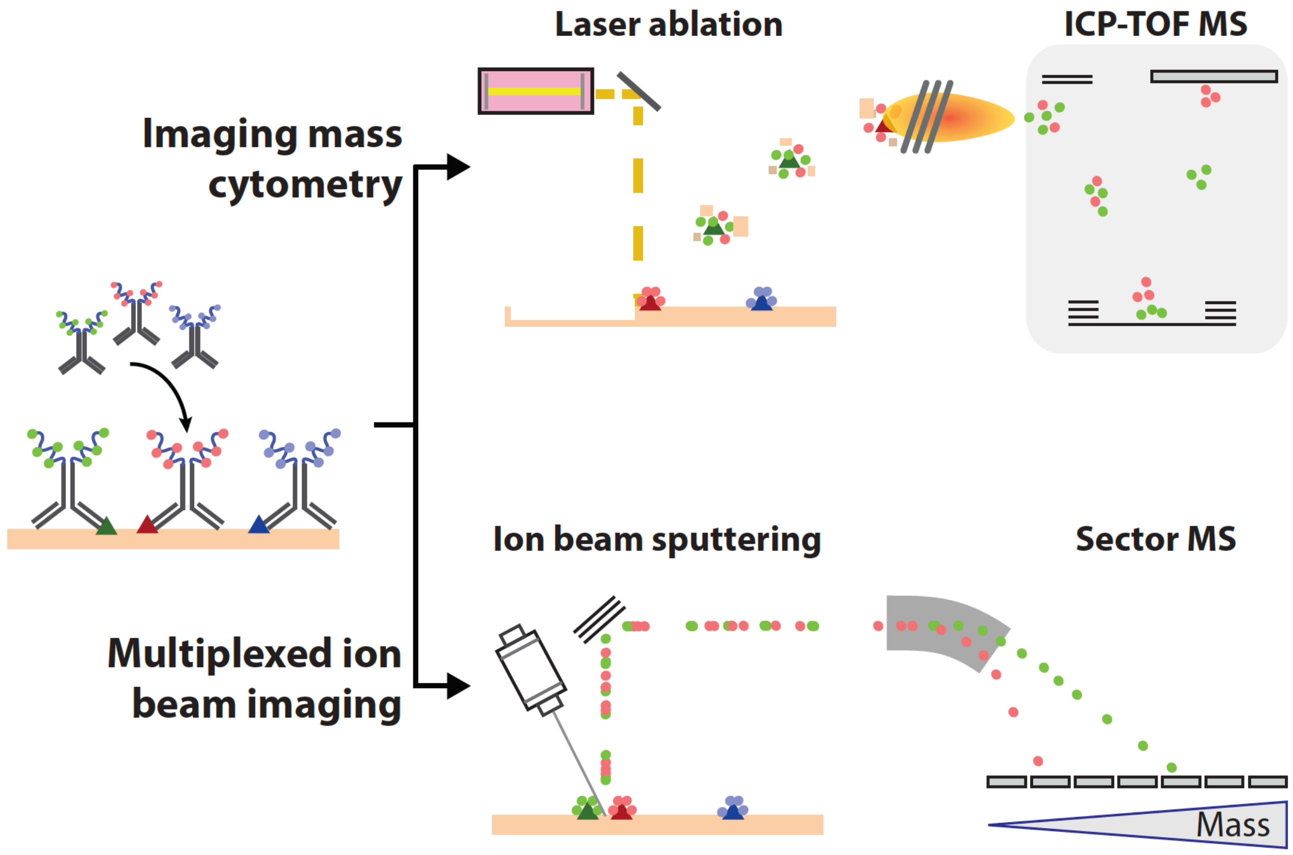
Cell type interaction

Marker spatial association

Cell localisation

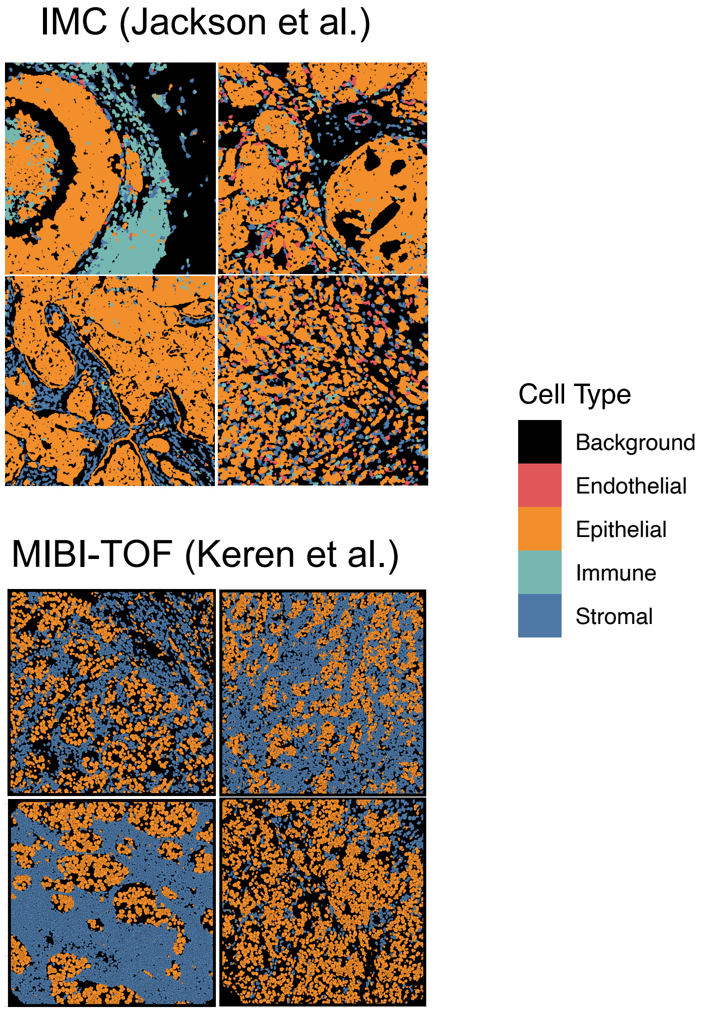
Junttila, M. R., & de Sauvage, F. J. (2013). Influence of tumour micro-environment heterogeneity on therapeutic response. *Nature*, *501*(7467), 346-354.

Data: Two Mass Cytometry Imaging data in breast cancer

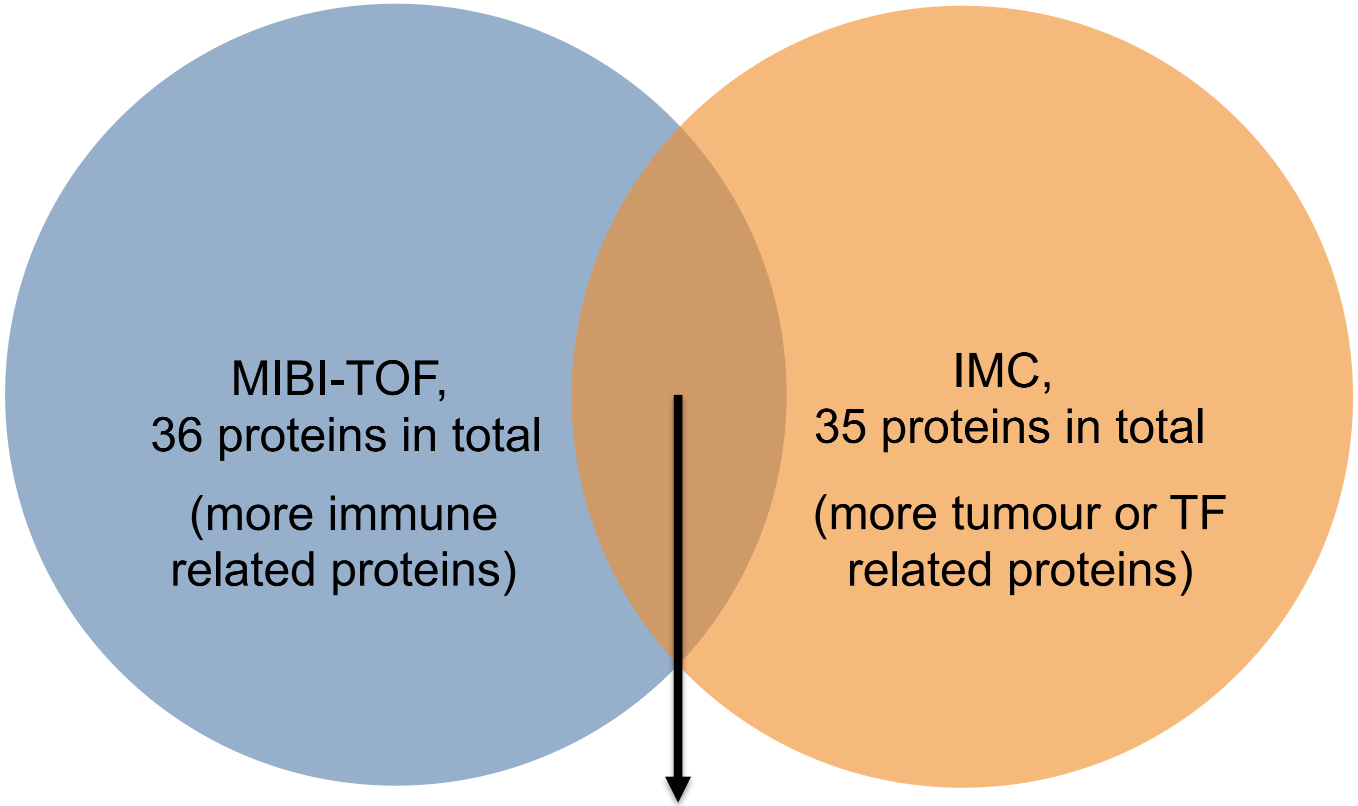


Bodenmiller, B. (2016). Multiplexed epitope-based tissue imaging for discovery and healthcare applications. *Cell systems*, 2(4), 225-238.

High dimensional images



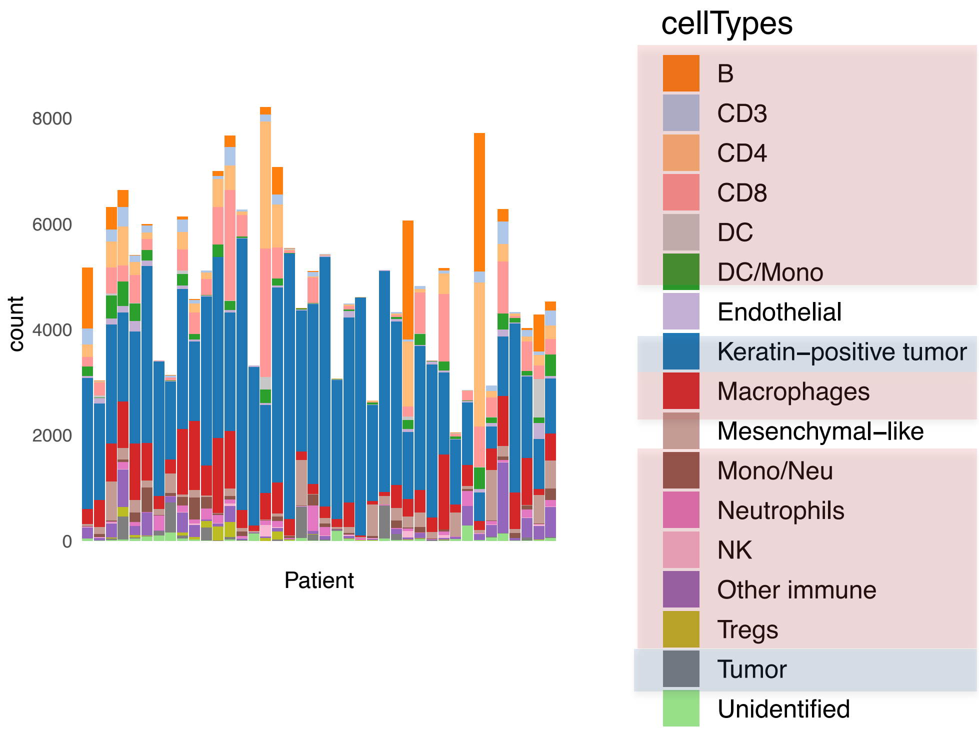
Integrative analysis challenge 1: Limited overlapped proteins



13 Common proteins:
EGFR, Ki67, SMA, Vimentin, p53, panCK, CD20, vWF, H3K27me3, CD45, CD68, CD3, pS6

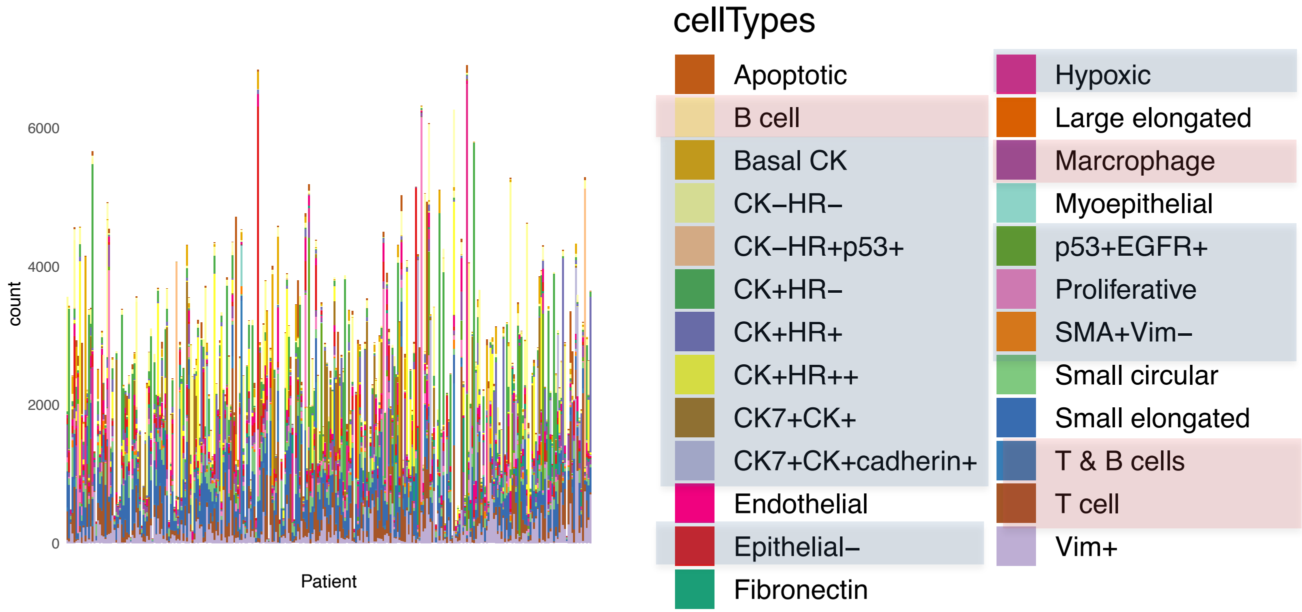
Integrative analysis challenge 2: Very different cell type annotation resolution

MIBI-TOF (Keren et al.)



Immune cell types

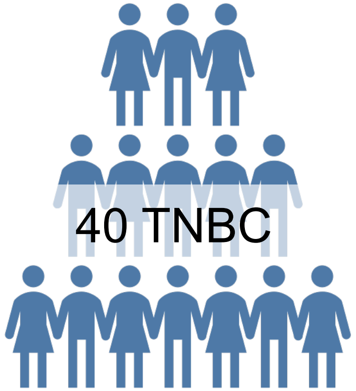
IMC (Jackson et al.)



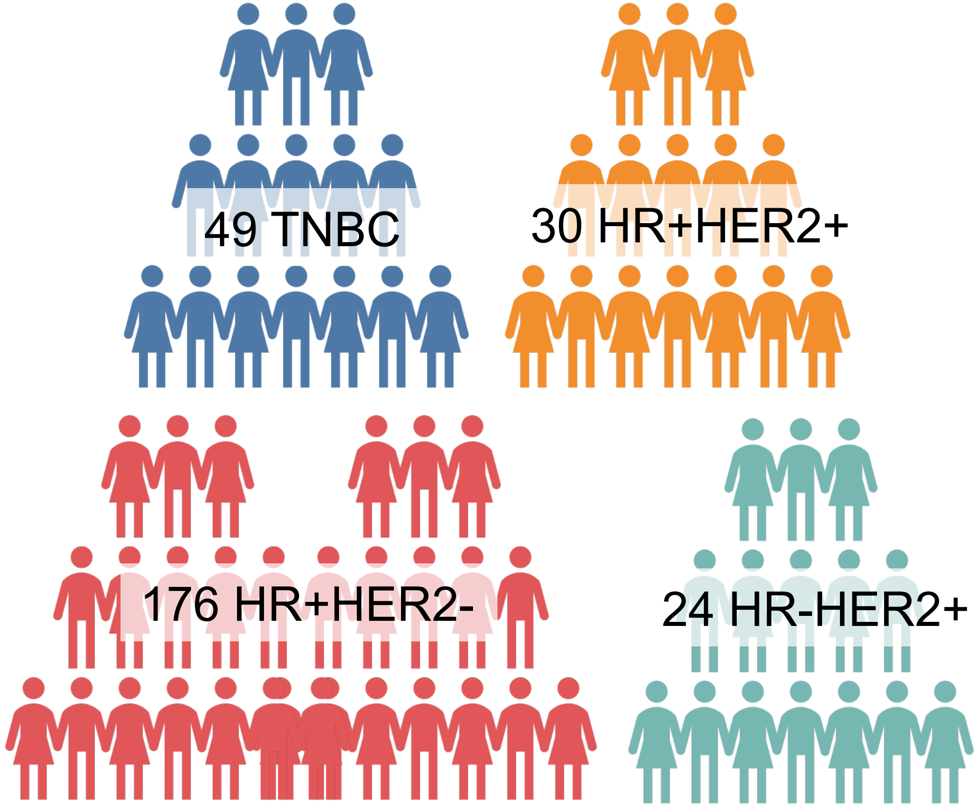
Epithelial cell types

Integrative analysis challenge 3: Partially overlapped of clinical types between these datasets

MIBI-TOF

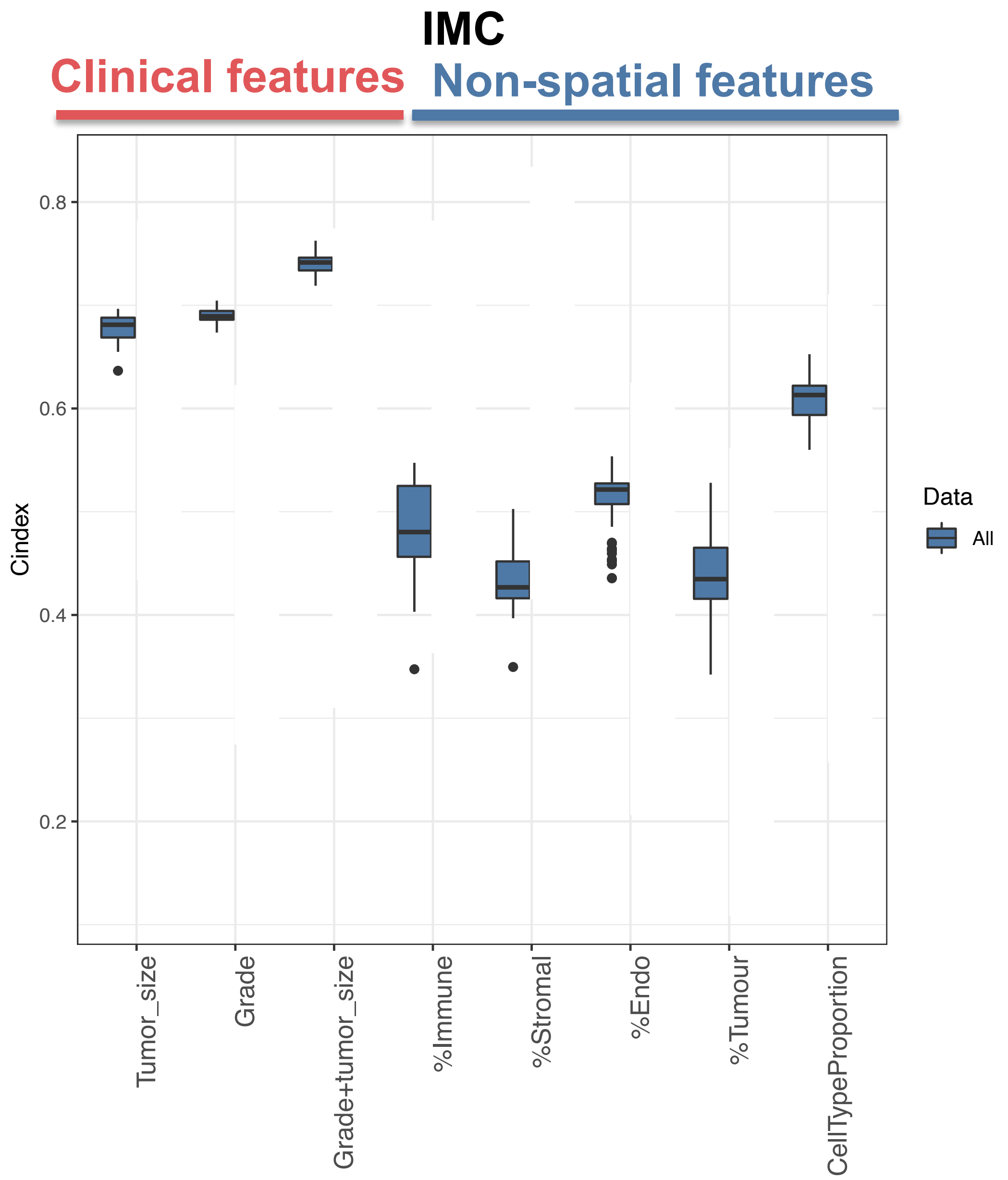


IMC



TNBC: Triple Negative Breast Cancer

Integrative analysis challenge 4: cohort heterogeneity



- Clinical features are more predictive than features like cell type proportion
- Clinical features perform poorly in triple negative breast cancer

Challenges and Questions

LIMITED OVERLAPPED PROTEIN

DIVERSE CELL TYPE ANNOTATION

PARTIALLY OVERLAPPED CLINICAL TYPE

HETEROGENEOUS COHORT

Spatial features

- Are the spatial features extracted from the images predictive in patient's survival?

Integrative analysis

- While integrating the two imaging data in the matrix level is challenging, are there common predictive spatial features shared between the two datasets?

Imputation

- Will other type of single-cell data information further improve the prediction in patient’s survival?

Challenges and Questions

Spatial features	• Are the spatial features extracted from the images predictive in patient's survival?
Integrative analysis	• While integrating the two imaging data in the matrix level is challenging, are there common predictive spatial features shared between the two datasets?
Imputation	• Will other type of single-cell data information further improve the prediction in patient's survival?

Spatial discriminative features

Survival analysis

Integration with CyTOF

Spatial features

Spatial discriminative features

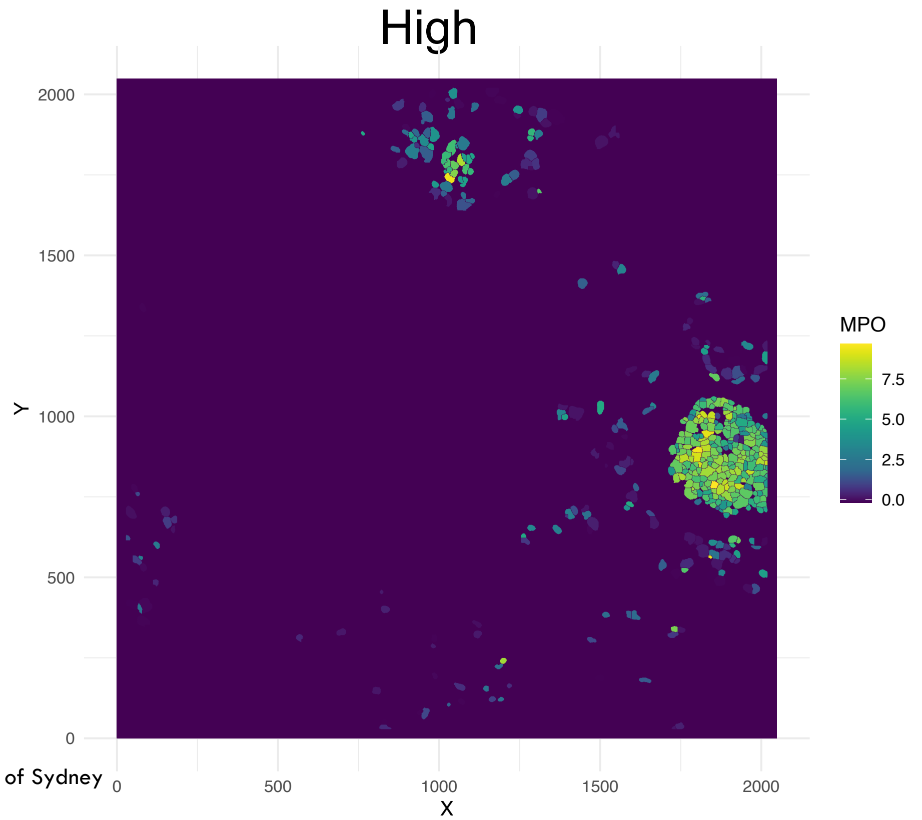
I. Moran's I	II. Nearest neighbour correlation	III. Cell type interaction composition	IV. L function

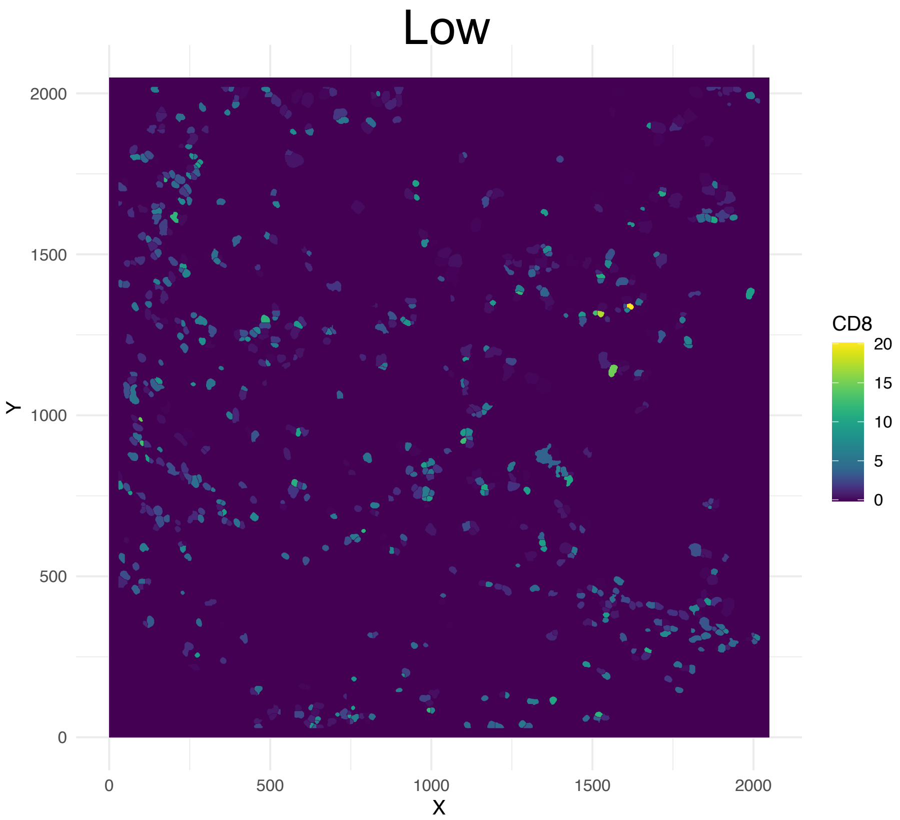
Spatial metrics I: Spatial autocorrelation - Moran's I

Moran's I measures spatial autocorrelation based on both feature locations and values simultaneously:

$$I = \frac{N}{\sum_{ij} w_{ij}} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

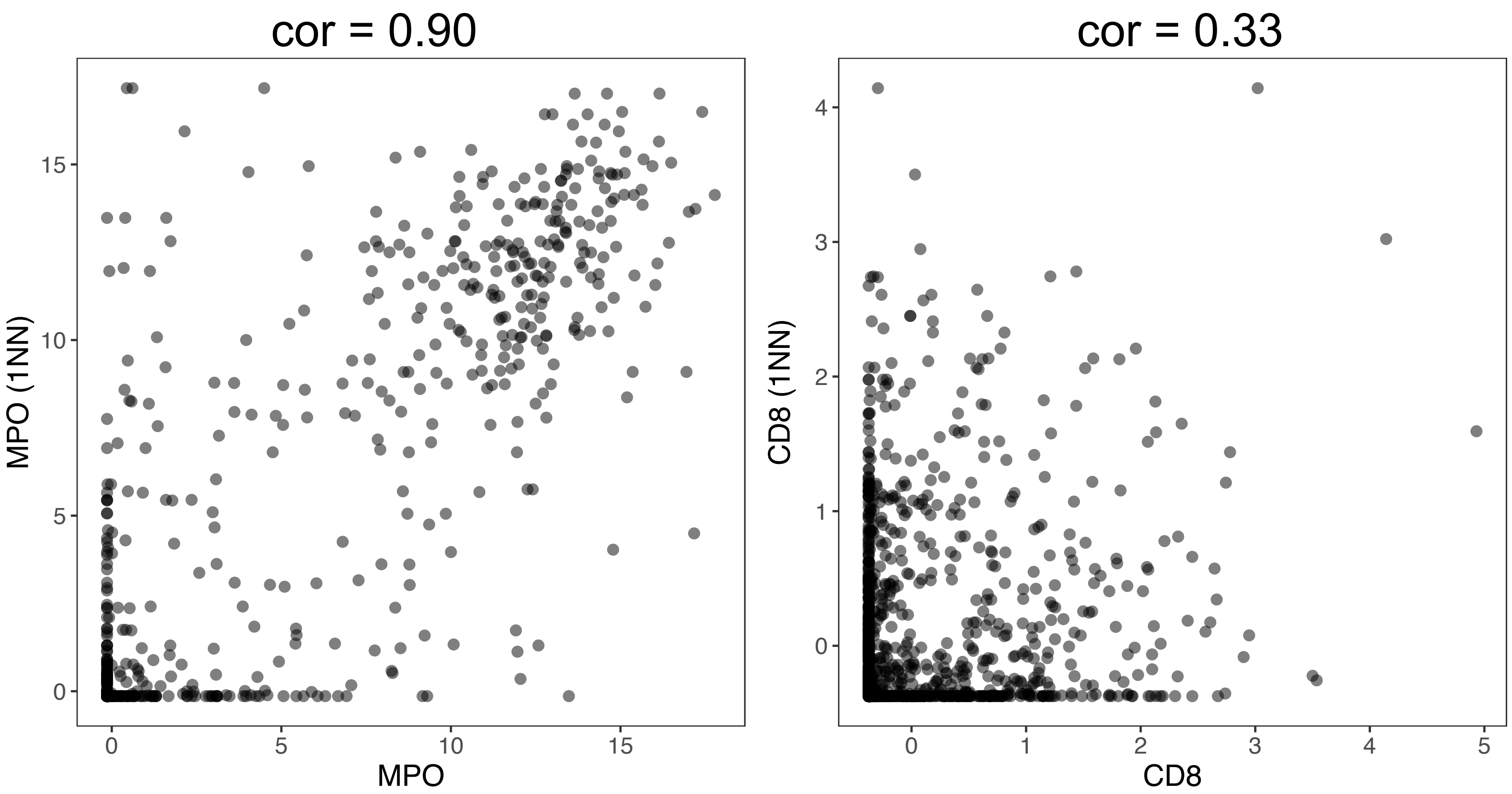
where N is the total number of cells indexed by i and j;
x is one epitope expression;
w_{ij} is a matrix of spatial weights





Spatial metrics II: Nearest Neighbour Correlation

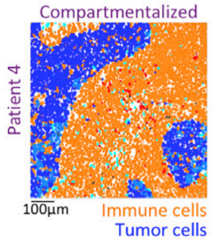
Nearest Neighbour Correlation: Correlation of protein expression between of cells with their nearest neighbours

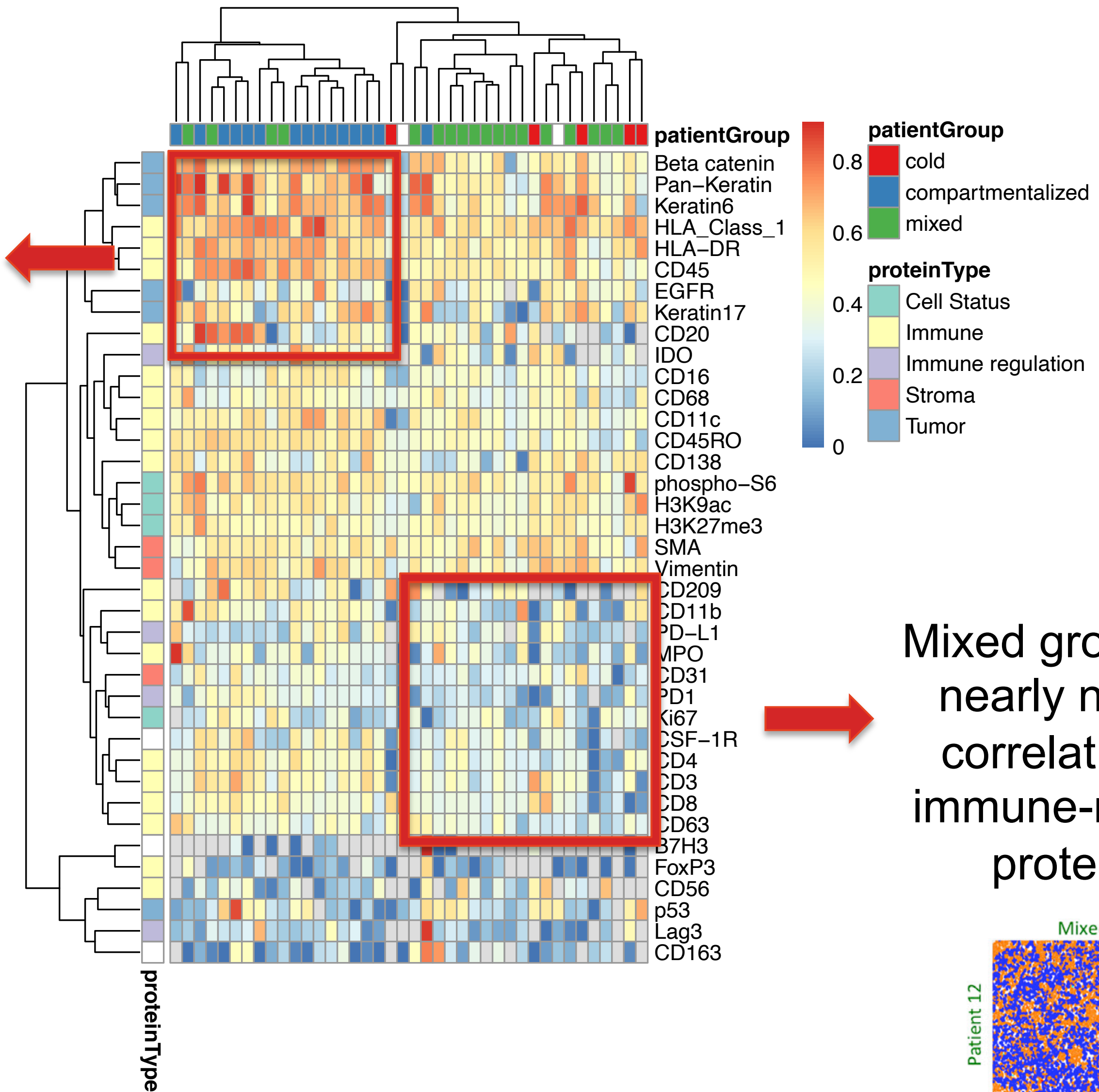


Spatial metrics II: Nearest Neighbour Correlation

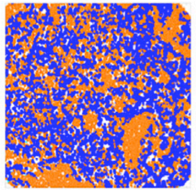
MIBI-TOF (Keren et al.)

Compartmentalized group has high NN correlation in tumor and immune related proteins



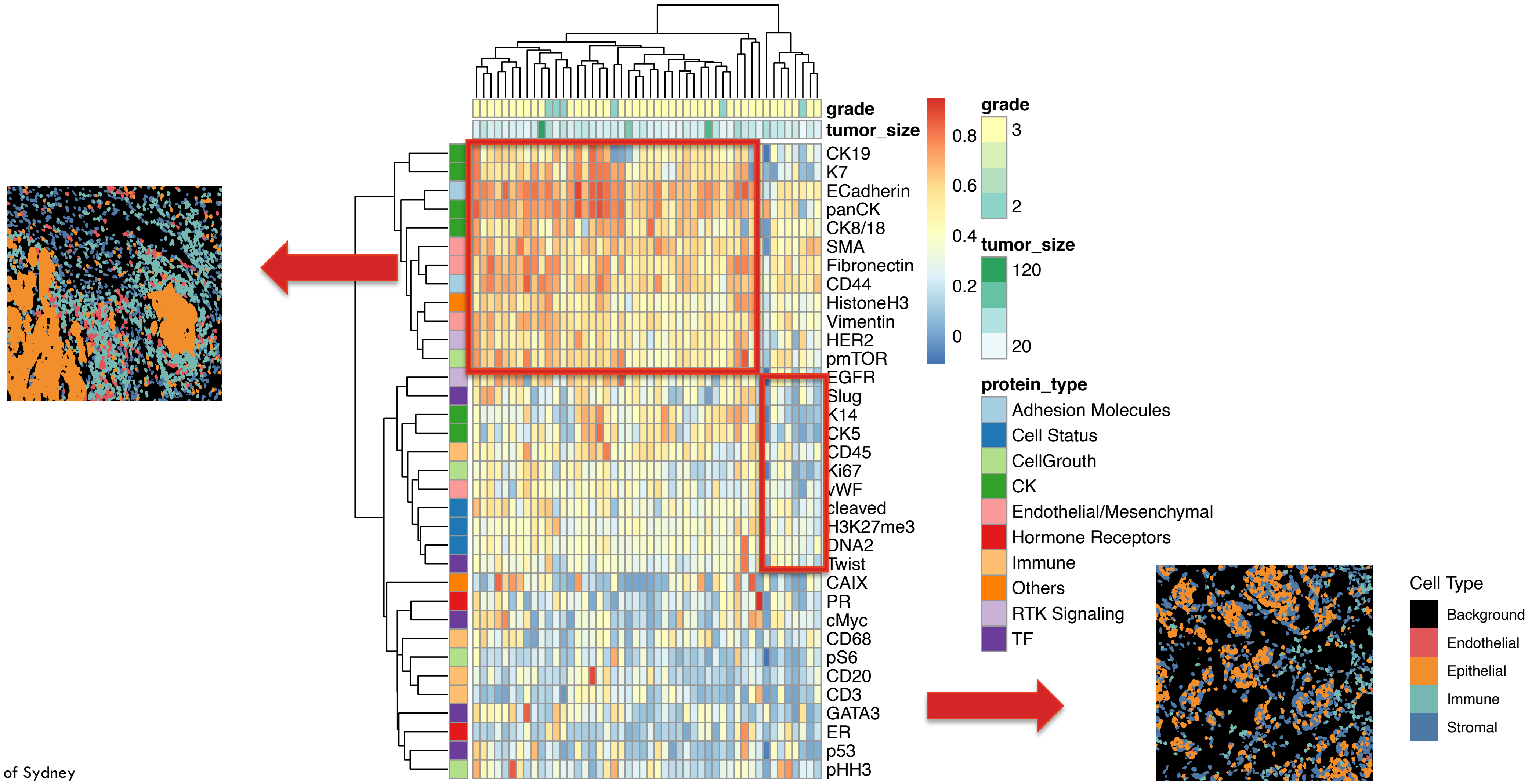


Mixed group has nearly no NN correlation in immune-related proteins

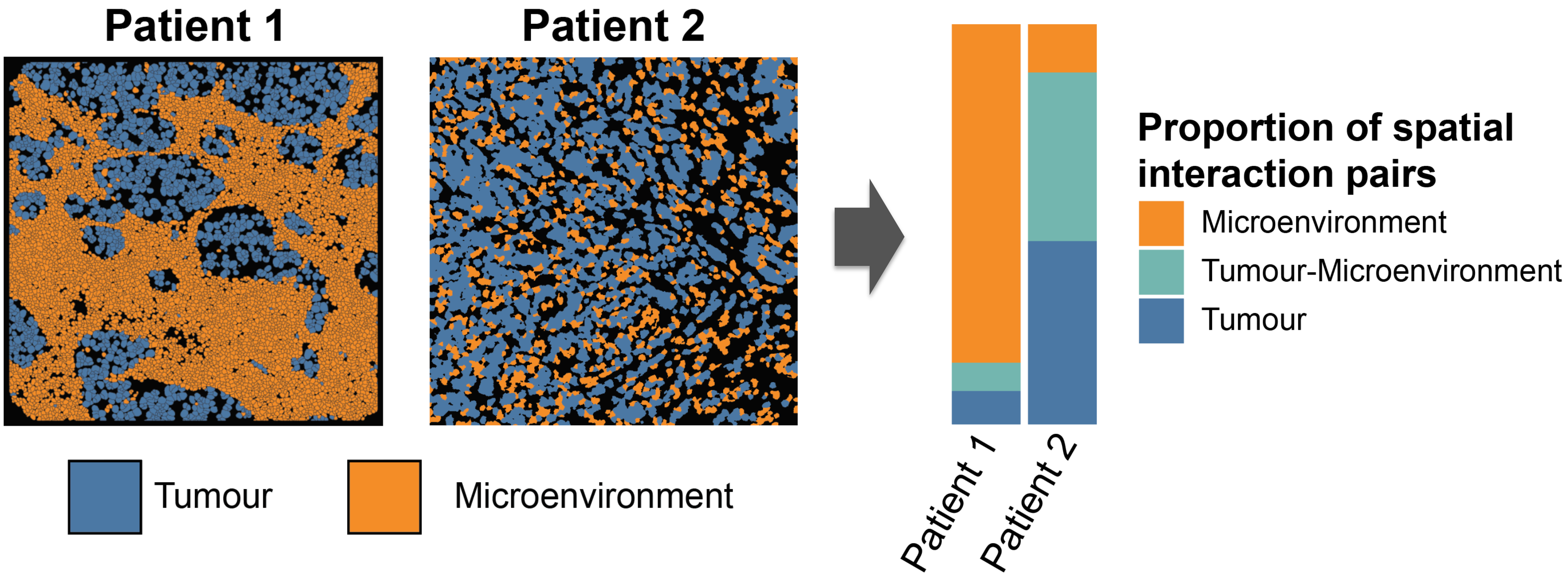


Spatial metrics II: Nearest Neighbour Correlation

IMC (Jackson et al.)

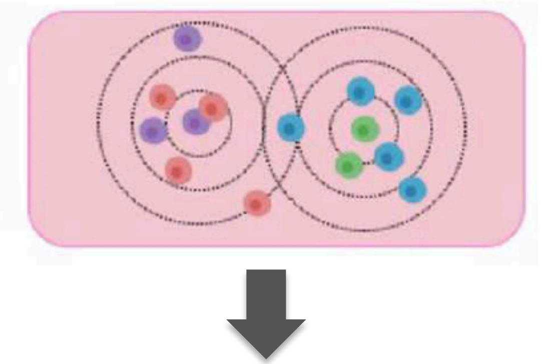


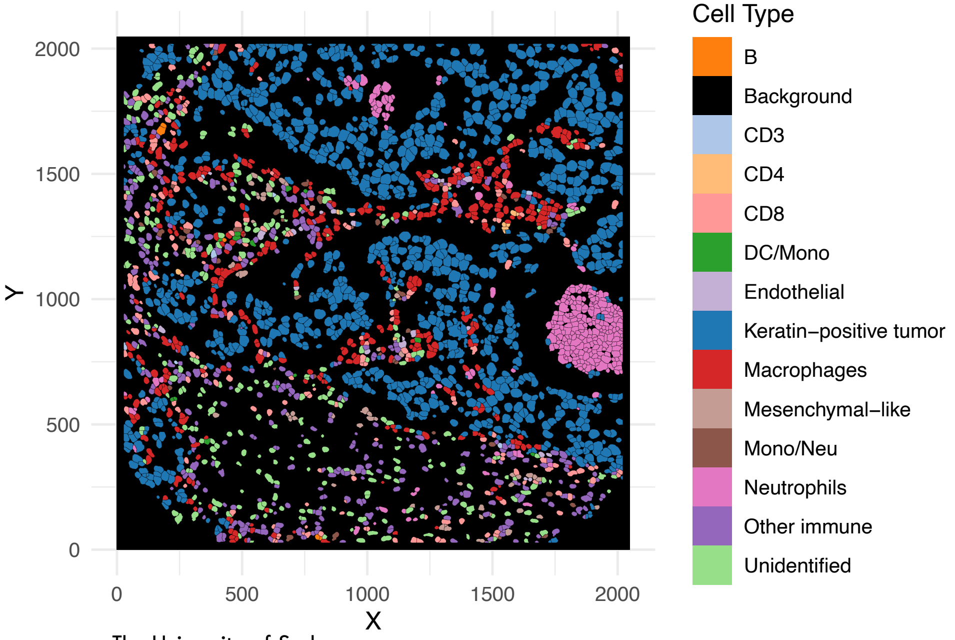
Spatial metrics III: Cell type interaction composition

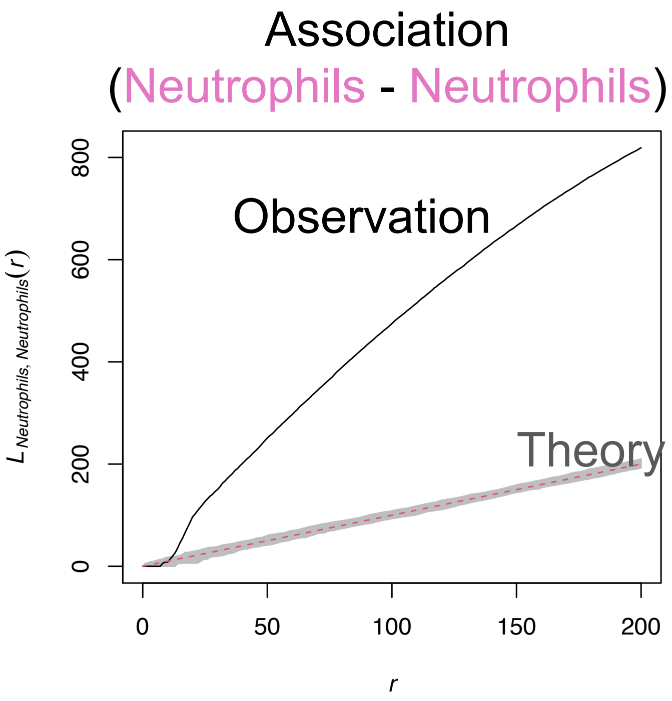


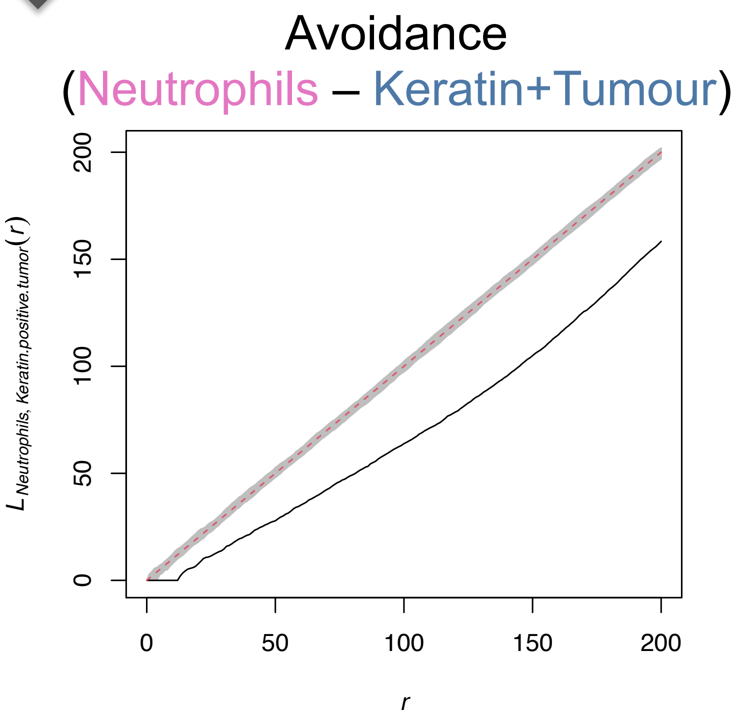
Spatial metrics IV: L functions

L functions: to assess the significance of cell-cell interactions









Baharlou, H., Canete, N. P., Cunningham, A. L., Harman, A. N., & Patrick, E. (2019). Mass Cytometry Imaging for the Study of Human Diseases—Applications and Data Analysis Strategies. *Frontiers in Immunology*, *10*.

Survival analysis using spatial features

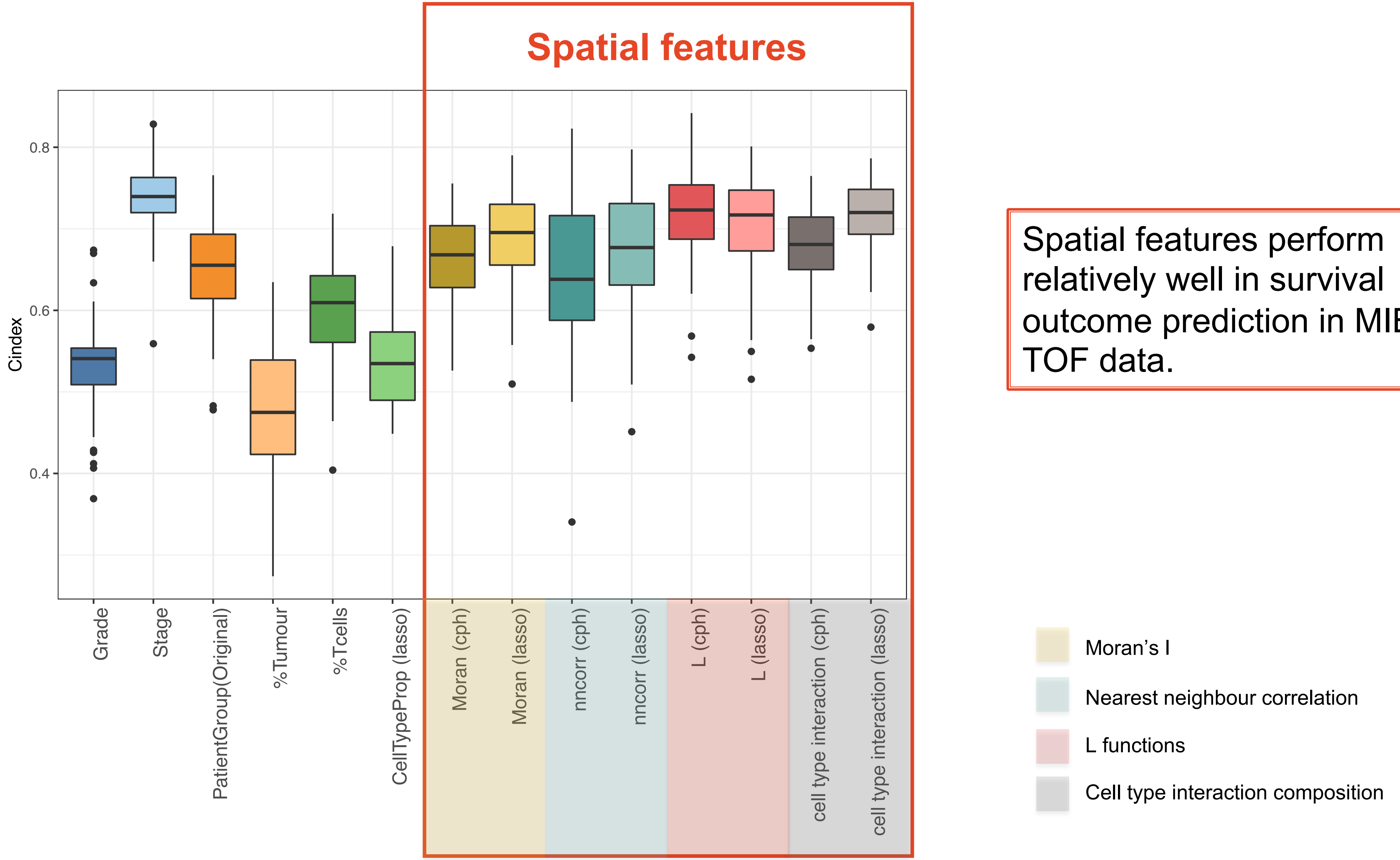
Spatial discriminative features

- I. Moran's I
- II. Nearest Neighbour correlation
- III. Cell type interaction composition
- IV. L function

Survival analysis

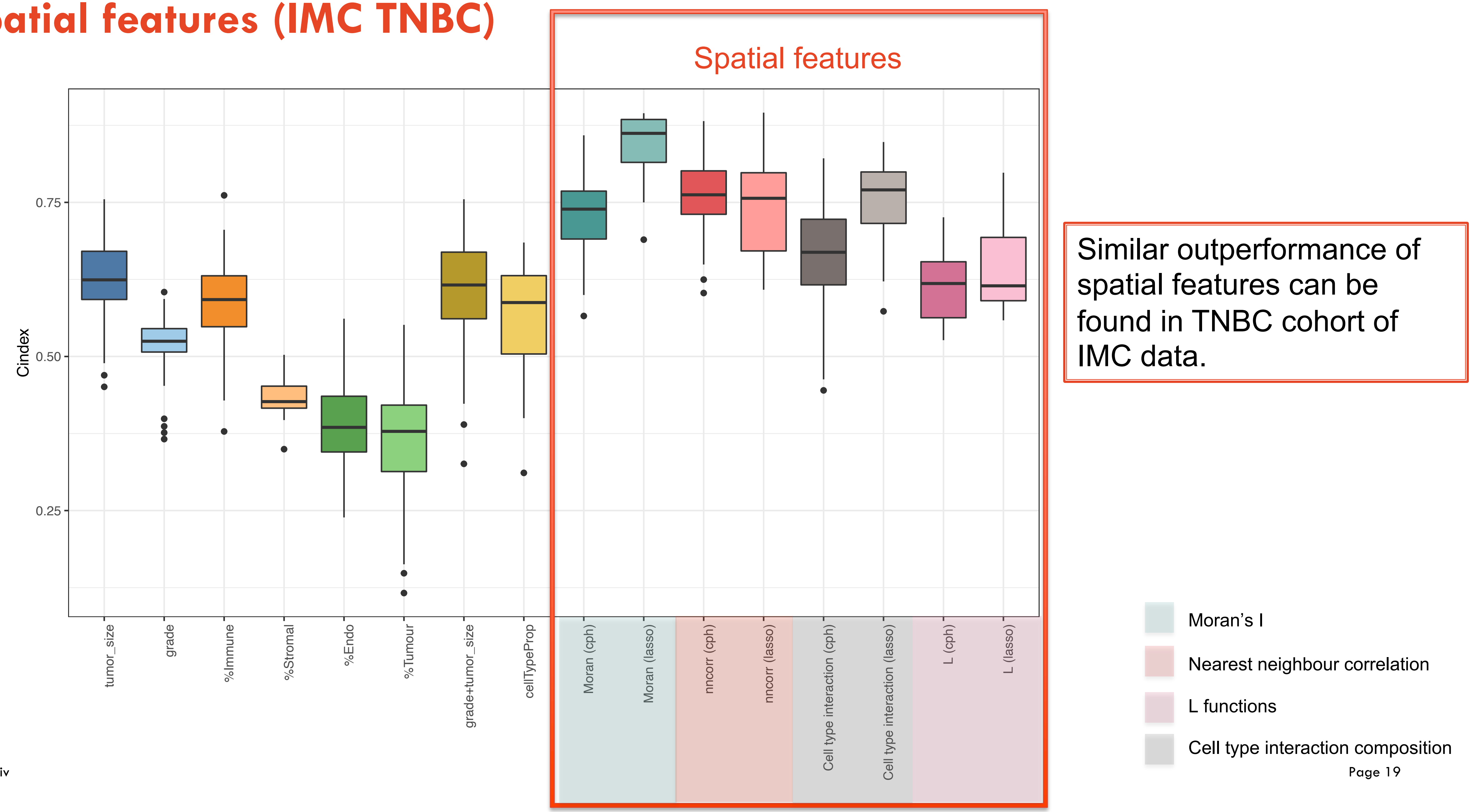
- Select features based on univariate cox model
- Fitting the multivariate cox model (coxph, lasso)
- Evaluating survival model using c-index via cross validation

Survival analysis (MIBI-TOF)



Spatial features perform relatively well in survival outcome prediction in MIBI-TOF data.

Spatial features (IMC TNBC)



Similar outperformance of spatial features can be found in TNBC cohort of IMC data.

Moran's I

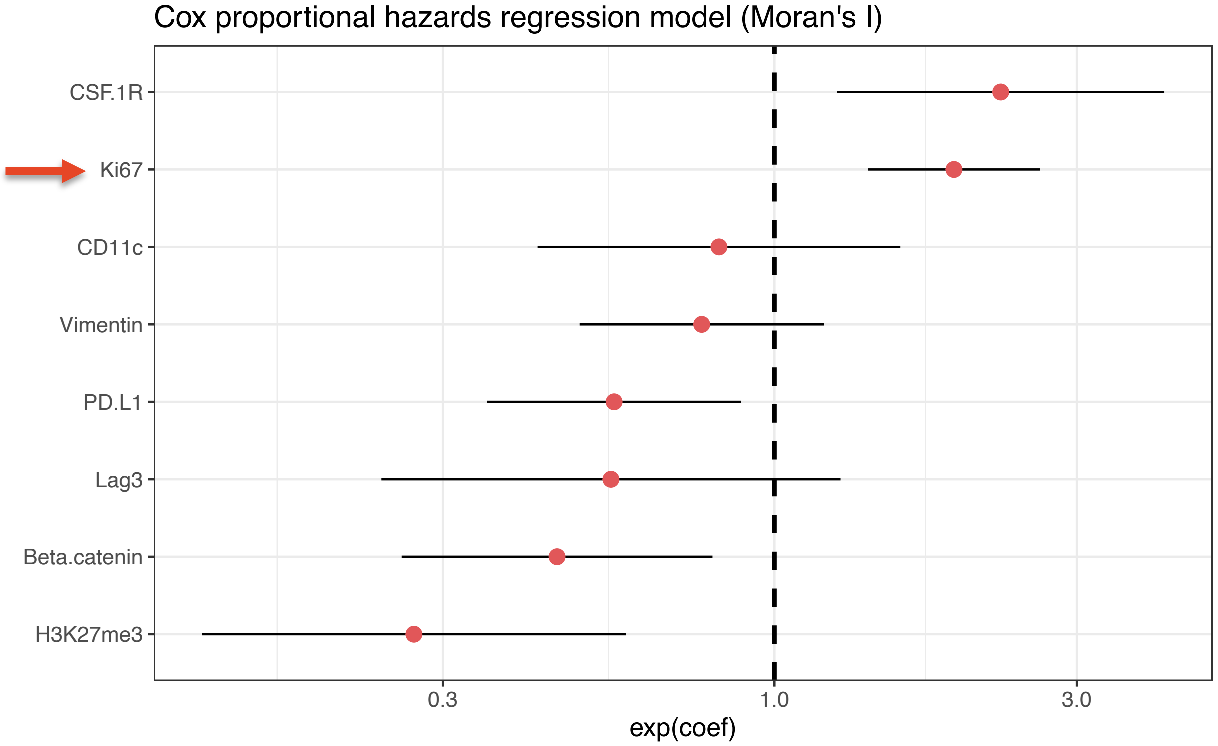
Mesenchymal markers
Immune markers
Cell status markers
Tumour markers
Immune regulation markers

MIBI-TOF | **IMC**

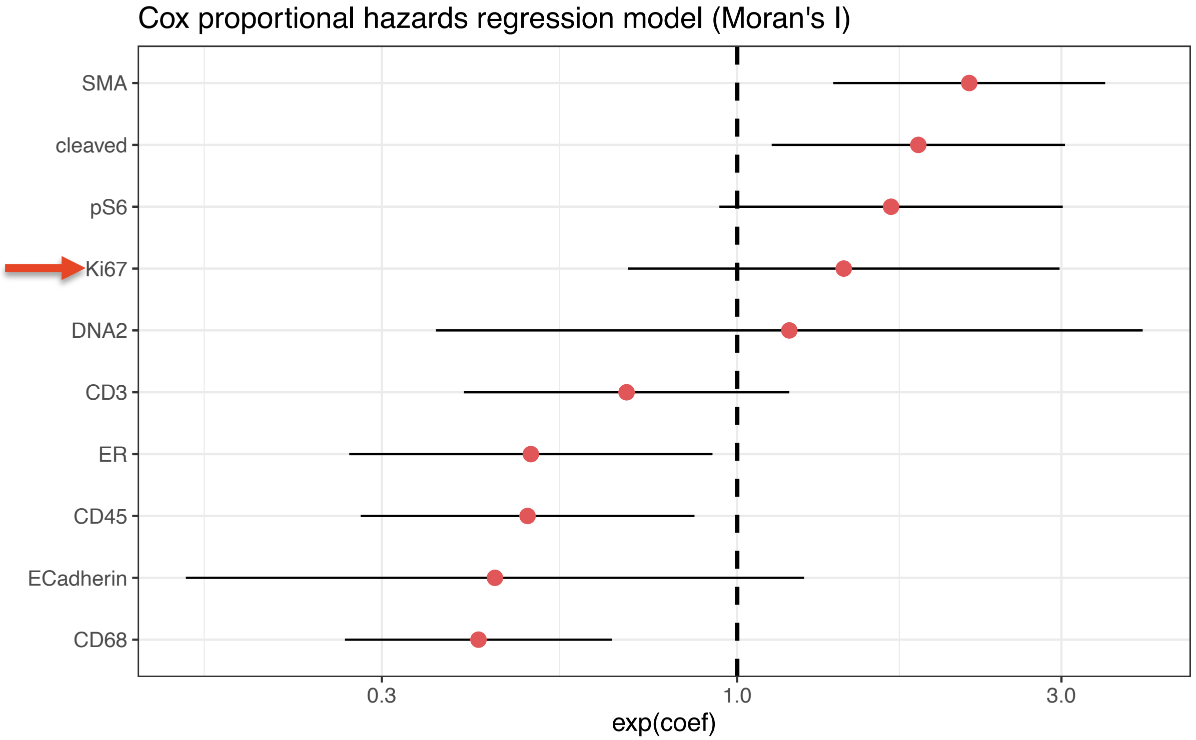
Top features are from diverse marker categories.

Moran’s I

MIBI-TOF



IMC



Only one common selected top features (Ki67, a cell proliferating marker) in two models

Optimal transport to impute imaging data using CyTOF data

Spatial discriminative features

I. Moran's I	II. Nearest Neighbour correlation	III. Cell type interaction composition	IV. L function

Survival analysis

Select features based on univariate cox model	Fitting the multivariate cox model (coxph, lasso)	Evaluating survival model using c-index

Integration with CyTOF

Optimal transport to predict additional protein expression

THE UNIVERSITY OF SYDNEY

Optimal transport

Optimal transport problem setting

$$\arg\min_{\gamma} < \gamma, M >_F + \lambda \sum_{i,j} \gamma_{ij} \log(\gamma_{ij})$$

s.t.

$$\sum_i \gamma_{ij} = a_j$$

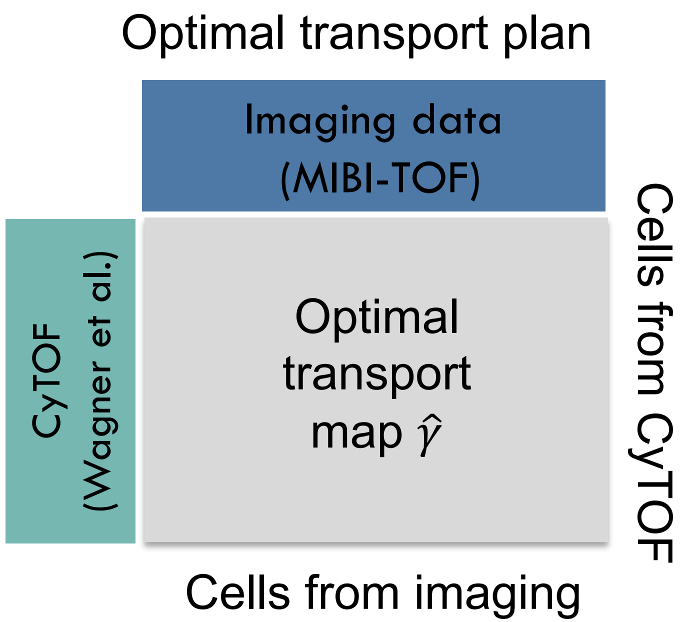
$$\sum_j \gamma_{ij} = b_i$$

$$\gamma \geq 0$$

where M is the cost matrix for the dissimilarity of cells between CyTOF and imaging data;
a_j is the weight for cell j in imaging data; and b_i is the weight for cell i in CyTOF data.

The prediction of a protein expression at position j in imaging using the CyTOF protein expression $g \in R^n$ is

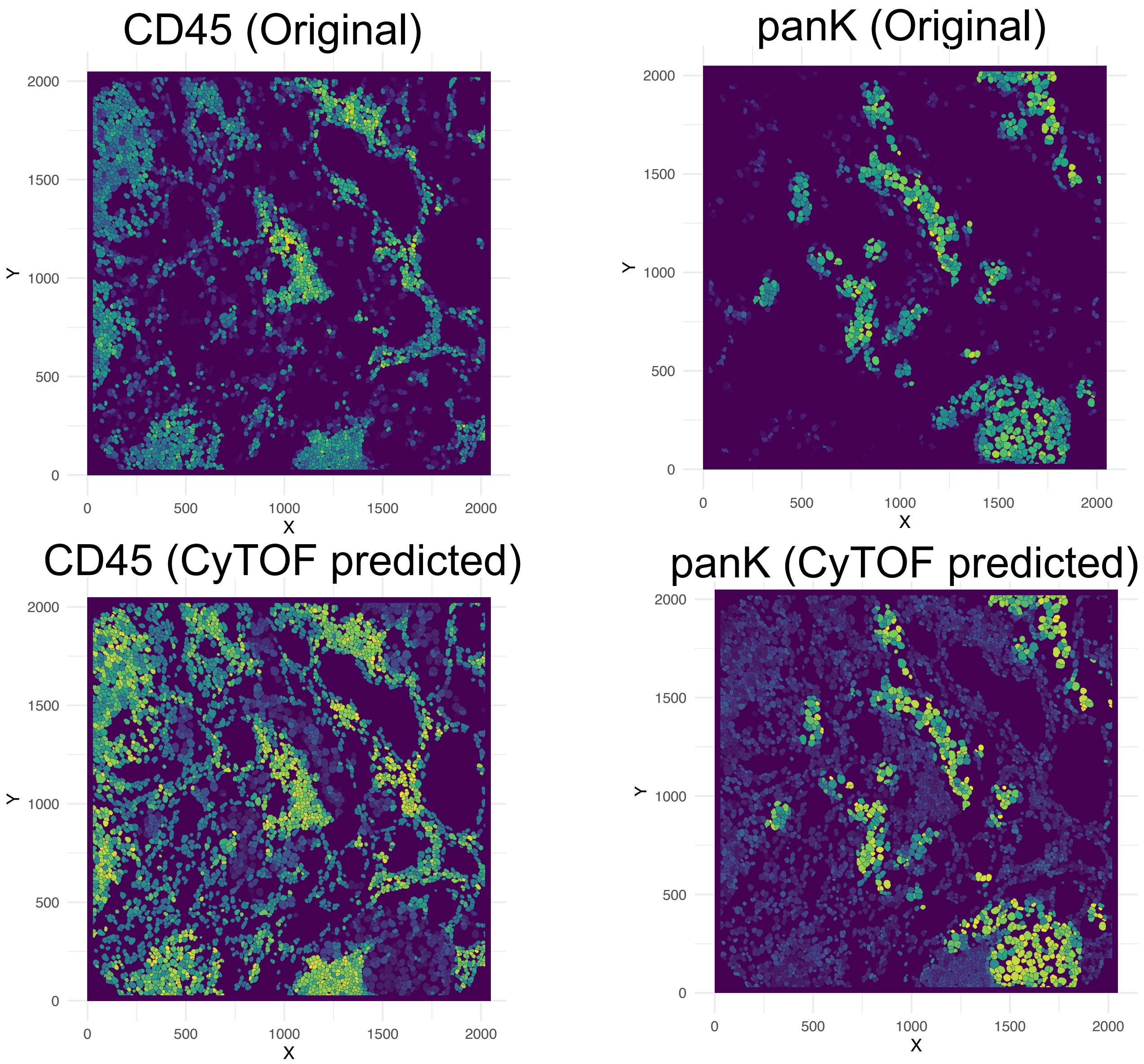
$$\frac{\sum_i \hat{\gamma}_{ij} \times g_i}{\sum_i \hat{\gamma}_{ij}}$$



Challenges:
- Wagner et al. CyTOF data only has ~14 proteins common with MIBI-TOF data

Cang et al. (2020) Nature Communication

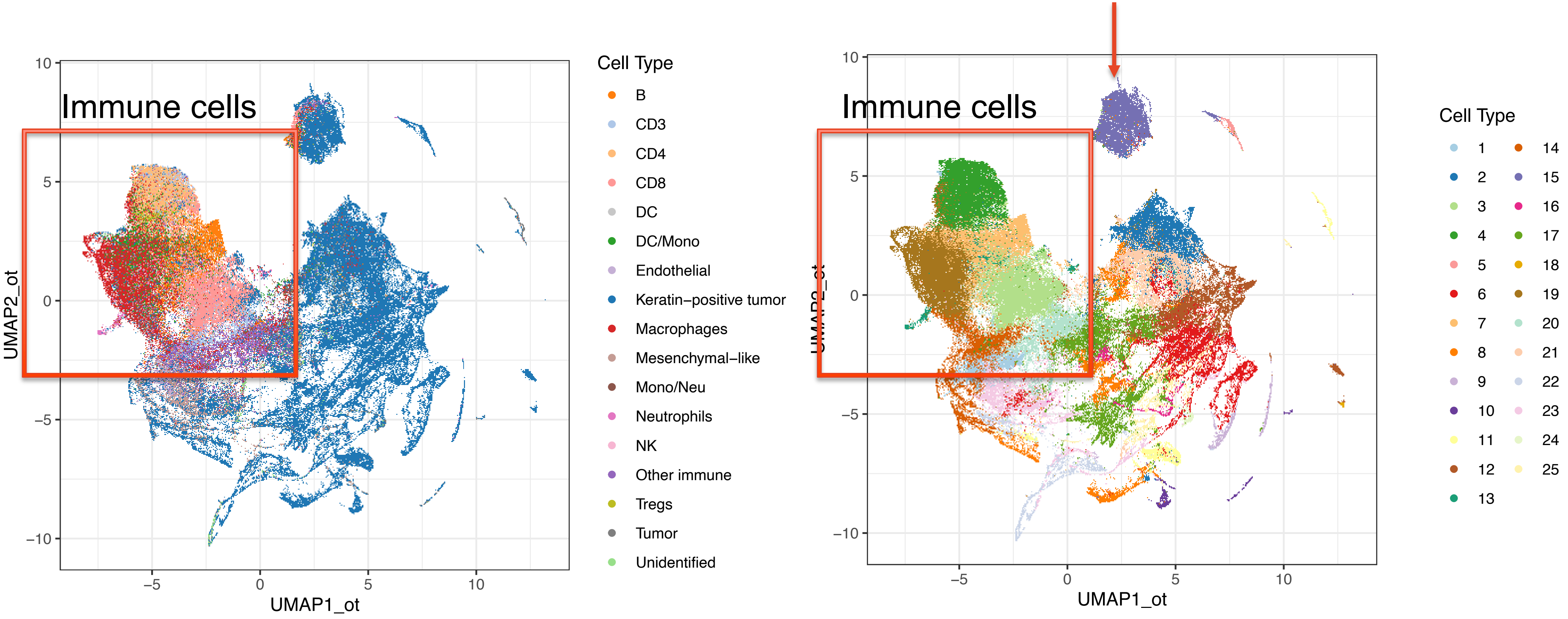
Protein prediction



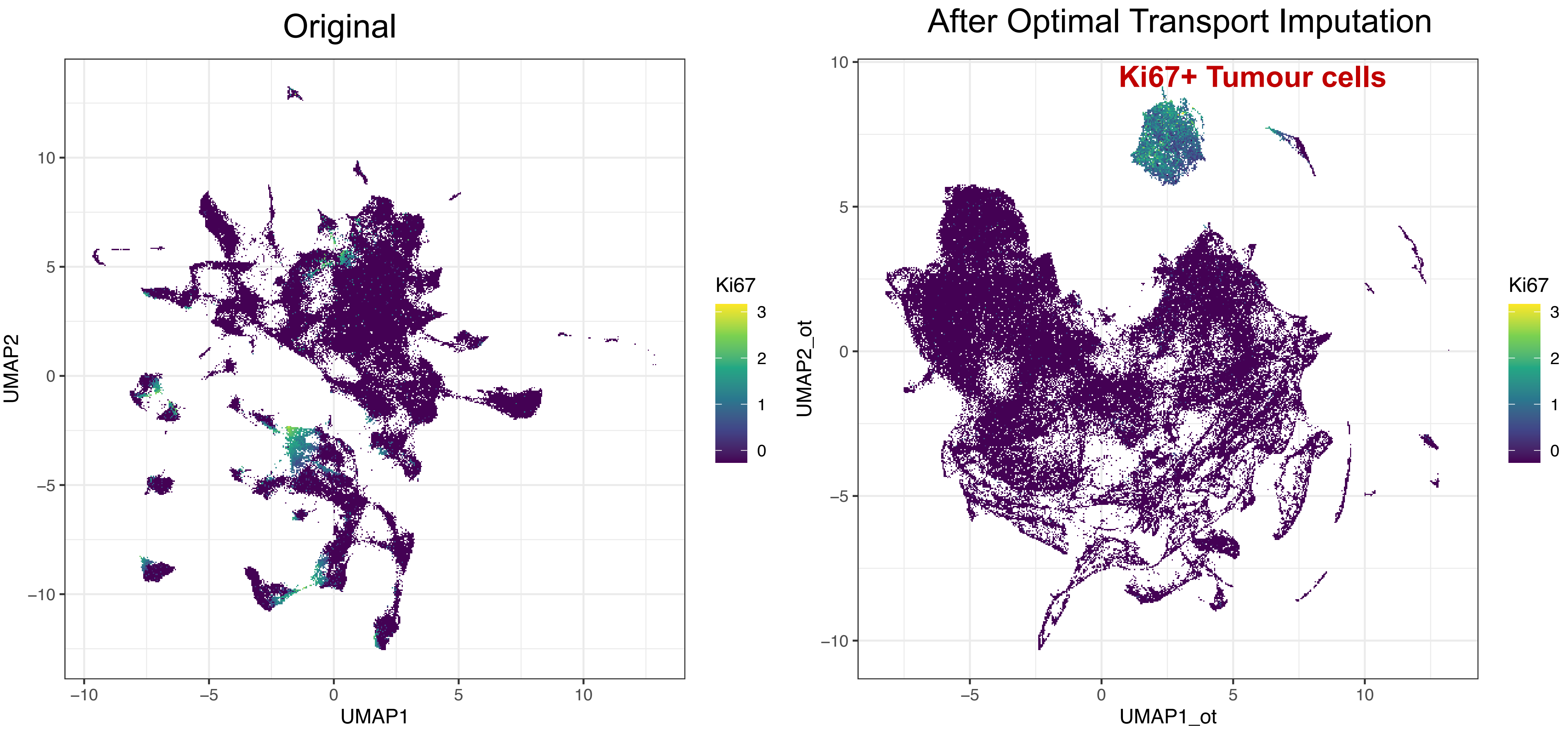
cor = 0.85

cor = 0.94

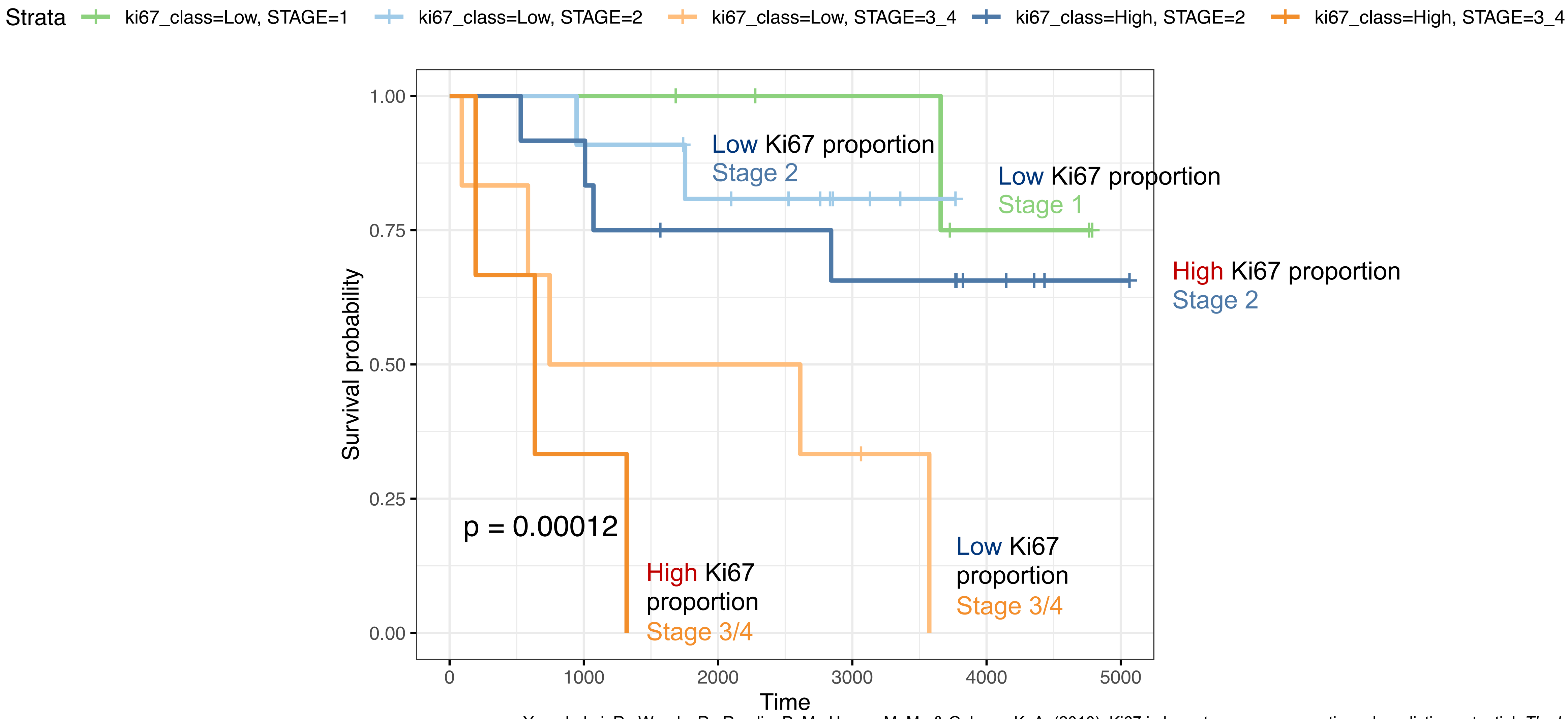
Identification of new sub cell type



Identification of new sub cell type



Survival analysis based on new annotations



Yerushalmi, R., Woods, R., Ravdin, P. M., Hayes, M. M., & Gelmon, K. A. (2010). Ki67 in breast cancer: prognostic and predictive potential. *The lancet oncology*, *11*(2), 174-183.

Inwald, E. C., Klinkhammer-Schalke, M., Hofstädter, F., Zeman, F., Koller, M., Gerstenhauer, M., & Ortmann, O. (2013). Ki-67 is a prognostic parameter in breast cancer patients: results of a large population-based cohort of a cancer registry. *Breast cancer research and treatment*, *139*(2), 539-552.

Conclusion

- Spatial features especially spatial association of the protein expression improve the survival prediction in both imaging datasets.
- Limited common feature is found between the survival models due to the limited common proteins measured in two datasets.
- Imputation of additional protein expression improved the cell type identification, and have potentials to increase the commonality between two datasets and further improve the survival prediction.

Acknowledgment

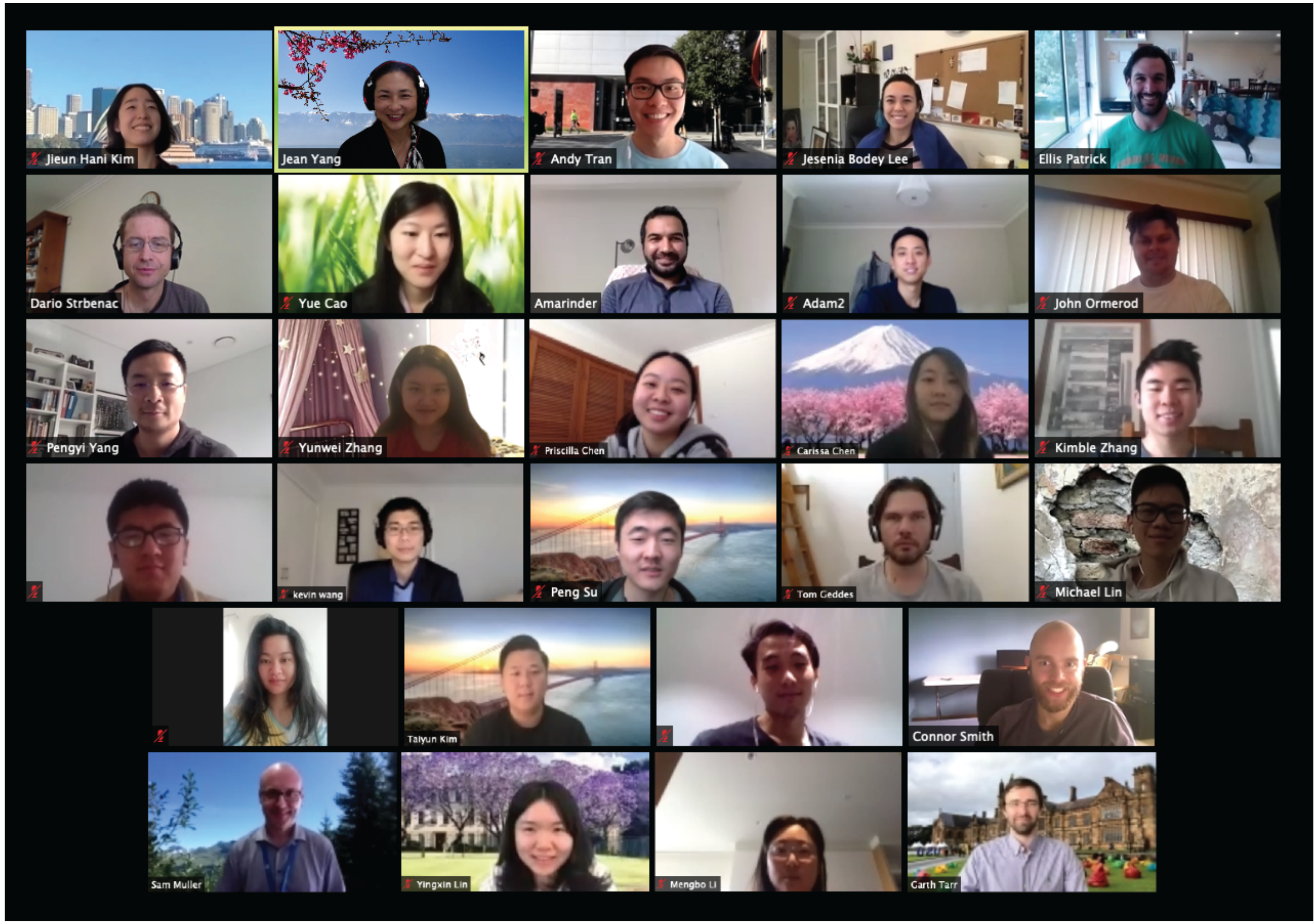
The University of Sydney

School of Mathematics and Statistics

Sydney Precision Bioinformatics Research Group

Jean Yang
Pengyi Yang
Samuel Mueller
John Ormerod
Rachel Wang
Ellis Patrick
Garth Tarr
Dario Strbenac

Yue Cao
Hani Kim
Thomas Geddes
Taiyun Kim
Mengbo Li
Connor Smith
Andy Tran
Andy Wang
Kevin Wang
Xiangnan Xu
Yunwei Zhang



Lauren Hsu and Aedin Culhane for providing the processed MIBI-TOF and CyTOF data