Enhancing Breast Cancer Survival Prediction Using Single-Cell Imaging Data

Aritra Ghosh[1], Freddy Quarshie[2], Colin Purtell[3] and Clara Wolff[2]

Indian Statistical Institute[1], University of Michigan[2], College of William and Mary[3]

Research Questions

- How can we **model the survival time of breast cancer patients** using clinical information?
- Can we use **mass cytometry biomarker data** to increase the predictive accuracy of the model?
- Can we leverage the spatial component of **single-cell imaging data** to identify additional **predictors** associated with patient survival?

Introduction

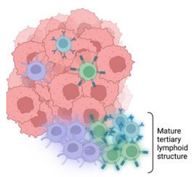
The Data

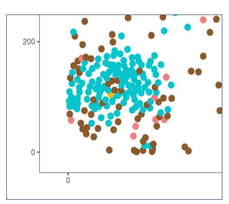
- 285 breast cancer patients from University Hospital Basel in Switzerland
- For each patient, a thin slice of the tissue around the tumor core was extracted and the expression levels of 33 biomarkers were measured for each cell using metal-tagged antibodies and a mass cytometry imaging system

Patient Information	Tumor Characteristics	Single-Cell Imaging
• Physical attributes (gender, age) • Survival time (months) • Disease stage	• Clinical type and grade • Presence of various cell classes • Tumor size (μm)	• Expression levels of 33 biomarkers • Location in tissue sample • Cell class

Identifying TLS from Single-Cell Data

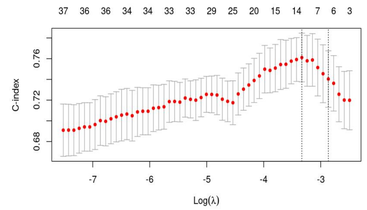
- Tertiary lymphoid structures (TLS) are often linked to more efficient immune response and increased survival time for certain cancer cases, however little is currently known about their role in breast cancer prognosis
- We were interested in whether we could use our single-cell imaging data to identify instances of TLS and incorporate our results into a survival model as a predictor





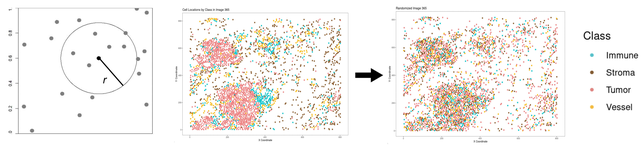
Methods

Survival Analysis: We used the Cox Proportional Hazards Regression to model survival time of breast cancer patients, the C-statistic to measure its accuracy, and LASSO to conduct variable selection

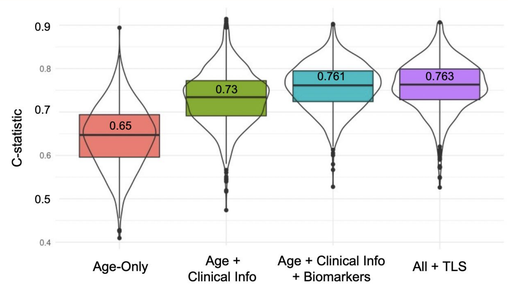


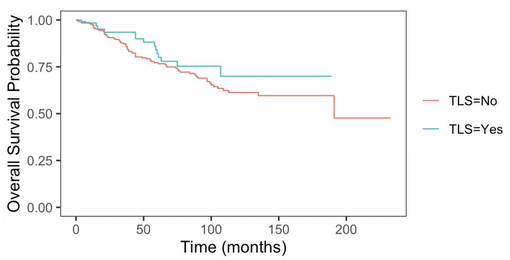
TLS Identification: We classified tissue samples containing potential TLS using Ripley's K-function paired with permutation tests to quantify abnormal clustering of immune cells in the spatial single-cell data

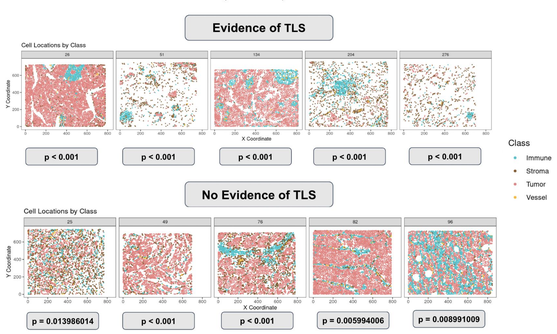
Ripley's K: $\widehat{K}(t) = \lambda^{-1} \sum_{i=1}^{n} \sum_{j \neq i} \frac{I(d_{ij} \leq r)}{n}$

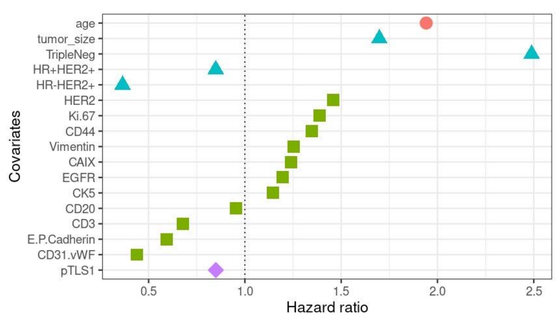


Results

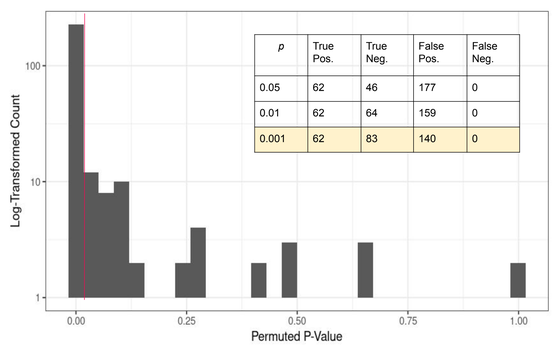








- Biomarker information increased predictive accuracy for breast cancer survival by 4%. Incorporating TLS presence raised it by an additional 0.3%
- In our final model, after incorporating the K-function classification, the TLS covariate had a hazard ratio of 0.85
- After a visual inspection to create a set of true positives and negatives, 62 images had evidence of TLS and 223 did not have evidence of TLS



p	True Pos.	True Neg.	False Pos.	False Neg.
0.05	62	46	177	0
0.01	62	64	159	0
0.001	62	83	140	0

Conclusions

Key Takeaways:

- Adding biomarker information to the standard prediction model of demographic and clinical information increased its predictive accuracy
- TLS predictor was associated with increased survival time
- Use of automated spatial analysis classification in identifying TLS produced promising results given the limited data
- We were able to successfully use single-cell imaging data to enhance the preliminary survival model and recognize statistically useful biological predictors. We hope this method can eventually be extrapolated to larger datasets

Limitations:

- Relatively small sample size of 285 patients makes effective validation of the predictive model difficult
- Lack of definitive cases of TLS presence to be used as true positives or training data
- Tissue samples that were imaged are only one cell thick, making the detection of three-dimensional biological clusters difficult
- Permuted K-function classification algorithm returned tissues with immune cell clustering that does not resemble a TLS

Future Work

- Acquire data containing multiple tissue samples from the same tumor core for a more comprehensive and layered image of cell locations and three-dimensional structures
- Conduct a more complete biological analysis with an imaging method that produces accurate classification of immune cell subtypes (T-cells and B-cells)
- Experimentation with AI image recognition models to reduce subjectivity of manual classification
- Implementation of a semiparametric AFT (Accelerated Failure Time) model in place of the Cox model

Acknowledgements

Thank you to Dr. Bhramar Mukherjee, Dr. Krithika Suresh, Dr. Matt Zawistowski, Dr. Xiang Zhou, Zheng Li, Elysia Chou, Dr. Maria Masotti, Dr. Phil Boonstra, Hanna Venera, Kirill Sabitov, Sabrina Olsson, and all BDSI staff and fellow participants

Reference

Jackson, H.W., Fischer, J.R., Zanotelli, V.R.T. et al. The single-cell pathology landscape of breast cancer. Nature 578, 615–620 (2020). https://doi.org/10.1038/s41586-019-1876-x