AI for Analysing Multiplexed Tissue Images

Shai Bagon[1] Yuval Bussi[2]
[1]*Weizmann AI Center (WAIC)*
[2]*Department of Molecular Cell Biology*

Abstract: Multiplexed imaging enables the measurement of multiple proteins in situ, offering an unprecedented opportunity to chart various cell types and states in tissues. These imaging modalities are penetrating every aspect of biological research in diverse fields, including cancer, immunology, development and neuroscience. However, analyzing these multiplexed tissue images is a challenging task. Current workflows are mostly based on pipelines that were developed for isolated cells in suspension and thus require manual inspection and annotation of multi-channel images, which is difficult and tedious. Most importantly, these methods do not scale to allow large-scale studies across many multiplexed images. Introducing AI to these workflows should give rise to faster and at-scale analysis of multiplexed tissue images.

Background. The tumor microenvironment (TME) is comprised of many different cell types, including malignant cells, but also non-transformed cells, such as fibroblasts, immune cells, neurons, lymphatics and vasculature. Each of these cell types can further assume a variety of phenotypes, defined by co-expression of multiple proteins. These cells organize in spatially-structured arrangements, exhibiting microenvironmental niches, nutrient gradients and cell-cell interactions. In recent years, it has become increasingly appreciated that this ecosystem, collectively termed the TME, plays a crucial role in mediating complex phenomena, such as tumor progression and response to treatment [1, 2].

Potential bottlenecks. Recent technological advancements in spatial profiling methodologies hold promise to fully capture the complexity of the TME and elucidate trans-cellular interactions in cancer. Methodologies like Multiplexed Ion Beam Imaging by Time of Flight (MIBI-TOF [3]) or Co-Detection by Indexing (CODEX [4]) allow gauging the expression of a multitude of proteins in tissue specimens while preserving tissue architecture. Multiplexed tissue images contain rich information on cell types and states and the sub-cellular localization of this data. However, analyzing these multiplexed images presents the biggest bottleneck to fully adopting these technologies in broader biological research. Analysis of this data is complex, and existing pipelines are limited and inadequate.

Analysis of multiplexed images is typically executed as a sequence of procedures, which includes image preprocessing [5], cell segmentation [6, 7], cell classification [8], and spatial analysis [9, 10]. These tasks are performed in a linear sequence, although it is clear that information from tasks that are executed later in the pipeline could be beneficial for tasks that are executed early. For example, knowledge of the classes of cells would highly improve cell segmentation. Moreover, these workflows often require manual inspection and annotation of multi-channel images, which is intractable due to the large number of proteins measured in each experiment. Finally, the workflows are often tailored to a specific panel of proteins or target tissue, and as such do not allow cross-study data integration. While artificial intelligence (AI) has been applied to tackle specific tasks in these workflows, its application has been very restrictive. Although some success is reported in tackling specific sub-tasks, e.g., cell segmentation [6, 7] and cell classification [11], curating the appropriate data and manually annotating it for each sub-task is still prohibitively time consuming and inefficient.

Altogether, the analysis of multiplexed images relies on fragmented pipelines with limited use of AI, mainly in the form of supervised models that require arduous labeling efforts.

Approach. In this project, we propose to develop a unified AI model for multiplexed tissue data. Inspired by recent trends in computer vision and deep learning, we take a more holistic approach: developing and training a single *representation* for the domain of multiplexed tissue images. This representation is an embedding of local multiplexed image regions (e.g., cells or TMEs) into high-dimensional vectors. Training this representation results in these vectors being organized in a *semantic* way in the high-dimensional embedding space. That is, we can expect vectors representing similar cells, or having similar biological properties, to be close to each other in the embedding space. Consequently, reasoning about biological properties becomes very trivial using this embedding. For example, one can perform cell classification by running simple algorithms, like k-means or k-nearest-neighbors, on the embeddings of cells to discover cell types.

To train this model with minimal manual annotation effort, we will also leverage new advances in self-supervised training methods, e.g. [12]. The efficacy of these self-supervised representations was recently demonstrated by tackling complex visual tasks, such as classification and segmentation, using simple computational tools applied to vectors in the learned embedding space [13, 14, 15].

Expected outcome. In this project, we will develop a self-supervised paradigm for training deep neural networks on multiplexed tissue images. The effectiveness of the learned representations will be showcased by tackling biological analysis tasks in a zero-shot manner. For example, demonstrating cell classification using a simple k-means clustering method, similar to [13]. If successful, this approach could revolutionize the field of multiplexed imaging analysis, making it widely accessible to cancer researchers and applicable to large clinical studies.

References

[1] M.-Z. Jin and W.-L. Jin, "The updated landscape of tumor microenvironment and drug repurposing," *Signal transduction and targeted therapy*, vol. 5, no. 1, pp. 1–16, 2020.

[2] X. Lei, Y. Lei, J.-K. Li, W.-X. Du, R.-G. Li, J. Yang, J. Li, F. Li, and H.-B. Tan, "Immune cells within the tumor microenvironment: biological functions and roles in cancer immunotherapy," *Cancer letters*, vol. 470, pp. 126–133, 2020.

[3] L. Keren, M. Bosse, S. Thompson, T. Risom, K. Vijayaragavan, E. McCaffrey, D. Marquez, R. Angoshtari, N. F. Greenwald, H. Fienberg, J. Wang, N. Kambham, D. Kirkwood, G. Nolan, T. J. Montine, S. J. Galli, R. West, S. C. Bendall, and M. Angelo, "Mibi-tof: A multiplexed imaging platform relates cellular phenotypes and tissue structure," *Science Advances*, vol. 5, no. 10, 2019.

[4] Y. Goltsev, N. Samusik, J. Kennedy-Darling, S. Bhate, M. Hale, G. Vazquez, S. Black, and G. P. Nolan, "Deep profiling of mouse splenic architecture with codex multiplexed imaging," *Cell*, vol. 174, no. 4, pp. 968–981, 2018.

[5] A. Baranski, I. Milo, S. Greenbaum, J.-P. Oliveria, D. Mrdjen, M. Angelo, and L. Keren, "Maui (mbi analysis user interface)—an image processing pipeline for multiplexed mass based imaging," *PLOS Computational Biology*, vol. 17, no. 4, pp. 1–16, 04 2021. [Online]. Available: https://doi.org/10.1371/journal.pcbi.1008887

[6] D. Bannon, E. Moen, M. Schwartz, E. Borba, T. Kudo, N. Greenwald, V. Vijayakumar, B. Chang, E. Pao, E. Osterman *et al.*, "DeepCell kiosk: scaling deep learning–enabled cellular image analysis with kubernetes," *Nature methods*, vol. 18, no. 1, pp. 43–45, 2021.

[7] C. Stringer, T. Wang, M. Michaelos, and M. Pachitariu, "Cellpose: a generalist algorithm for cellular segmentation," *Nature methods*, vol. 18, no. 1, pp. 100–106, 2021.

[8] M. J. Geuenich, J. Hou, S. Lee, S. Ayub, H. W. Jackson, and K. R. Campbell, "Automated assignment of cell identity from single-cell multiplexed imaging and proteomic data," *Cell Systems*, vol. 12, no. 12, pp. 1173–1186, 2021.

[9] Z. Wu, A. E. Trevino, E. Wu, K. Swanson, H. J. Kim, H. B. D'Angio, R. Preska, G. W. Charville, P. D. Dalerba, A. M. Egloff *et al.*, "Graph deep learning for the characterization of tumour microenvironments from spatial protein profiles in tissue specimens," *Nature Biomedical Engineering*, pp. 1–14, 2022.

[10] D. Schapiro, A. Sokolov, C. Yapp, Y.-A. Chen, J. L. Muhlich, J. Hess, A. L. Creason, A. J. Nirmal, G. J. Baker, M. K. Nariya *et al.*, "Mcmicro: A scalable, modular image-processing pipeline for multiplexed tissue imaging," *Nature methods*, vol. 19, no. 3, pp. 311–315, 2022.

[11] Y. Amitay, Y. Bussi, B. Feinstein, S. Bagon, I. Milo, and L. Keren, "CellSighter – a neural network to classify cells in highly multiplexed images," *bioRxiv*, 2022.

[12] M. Caron, H. Touvron, I. Misra, H. Jégou, J. Mairal, P. Bojanowski, and A. Joulin, "Emerging properties in self-supervised vision transformers," in *Proceedings of the International Conference on Computer Vision (ICCV)*, 2021.

[13] S. Amir, Y. Gandelsman, S. Bagon, and T. Dekel, "Deep ViT features as dense visual descriptors," in *What is Mostion for? Workshop, Europea Conference on Computer Vision (ECCVW)*, 2022.

[14] S. Vaze, K. Han, A. Vedaldi, and A. Zisserman, "Generalized category discovery," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 7492–7501.

[15] L. Melas-Kyriazi, C. Rupprecht, I. Laina, and A. Vedaldi, "Deep spectral methods: A surprisingly strong baseline for unsupervised semantic segmentation and localization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 8364–8375.

One-shot energy-based selection and optimization of diverse anti-tumor antibodies

Nachi Nathan (PI: Ziv Shulman), Ariel Tennenhouse (PI: Sarel Fleishman)

Monoclonal antibody (mAb)-based immunotherapy has become a chief component of cancer therapy along with surgery, chemotherapy, and radiation, with monoclonal antibodies having significant advantages over other methods due to the possibility of side-effect-free treatment[1]. One limiting factor of mAb treatment is the lack of speicific targets, as some cancers have no identifiable epitope that has been found. Therefore, sequencing tumor-infiltrating B cells is pivotal to understanding humoral tumor immunity and may lead to the discovery of antibodies against previously unknown targets. Recently, advances in single-cell RNA sequencing have enabled researchers to sequence the entire immunoglobulin (Ig) repertoire of B cells from the tumor microenviroment[2–7]. These experiments often result in hundreds to thousands of unique antibody sequences from which a small number must be selected to be evaluated as monoclonal antibodies. This selection is crucial as selected antibodies must display excellent "developability" characteristics (namely high stability, specificity, and expressibility)[8–10] to be useful therapeutics, and thus selecting the wrong antibody can lead to complete dead ends. **Here, we propose to leverage recent advances in machine learning-based protein structure prediction and computational antibody optimization[11,12] to develop a new method for selecting and optimizing antibody sequences based on structural and energetic principles[13–15].** This method will be completely automated, making it accessible for non-experts in antibody engineering. Ariel has experience engineering antibodies and has developed a method for antibody humanization, and Nachi has sequenced the B cells residing in the tumor-draining lymph nodes of several ovarian cancer patients and is now selecting mAbs to express, demonstrating their suitableness for this project. **This work can help to decipher the antibody-mediated response and identify novel therapeutics for ovarian cancer, which is poorly understood and does not respond to immunotherapy treatment**.

Current approaches rely on algorithms to generate clonal antibody trees based purely on antibody sequences, and a small number of members of the most clonally expanded B cells are selected, as they are predicted to have the highest affinity[16,17]. This approach is well suited for data from rodent immunizations with a strong clonal response leading to a small number of sequences to choose from but falls short for large and complicated datasets such as those obtained from single-cell RNA seq of anti-tumor antibodies. There may not be many expanded clones to choose from, and at some point, clonality becomes insignificant as the minimal differences provide no benefit[18]. In this way, current approaches will essentially select randomly from these expanded clones, which may not select for the most optimal antibodies and may limit the selection to antibodies against a target antigen (or epitope) that is common at a selected time point, thus reducing diversity.

Our method will first **eliminate any antibody sequences containing post-translational modification motifs**, which are known to cause antibody heterogeneity in therapeutic formulations. Then, we will predict the structure of each remaining antibody sequence acquired from the single-cell sequencing and cluster the predicted structures, yielding a set of diverse clusters. Within each cluster, we will use Rosetta all-atom calculations to rank the antibodies by energy and then use methods developed in the Fleishman Lab to optimize the stability, expressibility, and humanness of the top antibodies, yielding a small number of structurally diverse antibody sequences that are predicted to exhibit excellent developability characteristics[11,12]. Successful completion of this work will both allow for the **discovery of antibodies that traditional methods for selection would miss** in our data sets and will **remove the need for further optimization of discovered antibodies**.

References:

1. Zahavi, D. & Weiner, L. Monoclonal antibodies in cancer therapy. *Antibodies (Basel)* **9**, 34 (2020).
2. Bischoff, P. *et al.* Single-cell RNA sequencing reveals distinct tumor microenvironmental patterns in lung adenocarcinoma. *Oncogene* **40**, 6748–6758 (2021).
3. Paggetti, J., Seiffert, M. & Moussay, E. *New Insights into the Complexity of Tumor Immunology in B-cell Malignancies: Disease Biology and Signaling*. (Frontiers Media SA, 2022).
4. Wagner, J. *et al.* A single-cell atlas of the tumor and immune ecosystem of human breast cancer. *Cell* **177**, 1330–1345.e18 (2019).
5. Zhao, J. *et al.* Single-cell RNA sequencing reveals the heterogeneity of liver-resident immune cells in human. *Cell Discov.* **6**, 22 (2020).
6. Caligola, S., De Sanctis, F., Canè, S. & Ugel, S. Breaking the immune complexity of the tumor microenvironment using single-cell technologies. *Front. Genet.* **13**, 867880 (2022).
7. Thorsson, V. *et al.* The Immune Landscape of Cancer. *Immunity* **51**, 411–412 (2019).
8. Makowski, E. K., Wu, L., Desai, A. A. & Tessier, P. M. Highly sensitive detection of antibody nonspecific interactions using flow cytometry. *MAbs* **13**, 1951426 (2021).
9. Starr, C. G. *et al.* Ultradilute Measurements of Self-Association for the Identification of Antibodies with Favorable High-Concentration Solution Properties. *Mol. Pharm.* **18**, 2744–2753 (2021).
10. Lowe, D. *et al.* Aggregation, stability, and formulation of human antibody therapeutics. *Adv. Protein Chem. Struct. Biol.* **84**, 41–61 (2011).
11. Tennenhouse, A. *et al.* Reliable energy-based antibody humanization and stabilization. *bioRxiv* 2022.08.14.503891 (2022) doi:10.1101/2022.08.14.503891.
12. Warszawski, S. *et al.* Optimizing antibody affinity and stability by the automated design of the variable light-heavy chain interfaces. *PLoS Comput. Biol.* **15**, e1007207 (2019).
13. Evans, R. *et al.* Protein complex prediction with AlphaFold-Multimer. *bioRxiv* 2021.10.04.463034 (2022) doi:10.1101/2021.10.04.463034.
14. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589

(2021).

15. Abanades, B. *et al.* ImmuneBuilder: Deep-Learning models for predicting the structures of immune proteins. *bioRxiv* 2022.11.04.514231 (2022) doi:10.1101/2022.11.04.514231.
16. Yaari, G. & Kleinstein, S. H. Practical guidelines for B-cell receptor repertoire sequencing analysis. *Genome Med.* **7**, 121 (2015).
17. Glanville, J., Ponraj, P. & Ippolito, G. C. *Next-Generation Sequencing of Human Antibody Repertoires for Exploring B-cell Landscape, Antibody Discovery and Vaccine Development*. (Frontiers Media SA, 2020).
18. Mazor, R. D. *et al.* Tumor-reactive antibodies evolve from non-binding and autoreactive precursors. *Cell* **185**, 1208–1222.e21 (2022).