A Masked Image Modeling Approach to CyCIF Panel Reduction and Marker Imputation

Zachary Sims, Young Hwan Chang

Department of Biomedical Engineering and Computational Biology Program,
Oregon Health & Science University, Portland, Oregon, USA
{simsz, chanyo}@ohsu.edu

Abstract

Cyclic Immunofluorescence (CyCIF) has emerged as a powerful technique that can measure multiple biomarkers in a single tissue sample but it is limited in panel size due to technical issues and tissue loss. We develop a computational model that imputes a surrogate in silico high-plex CyCIF from only a few experimentally measured biomarkers by learning co-expression and morphological patterns at the single-cell level. The reduced panel is optimally designed to enable full reconstruction of an expanded marker panel that retains the information from the original panel necessary for downstream analysis. Using a masked image modeling approach based on the self-supervised training objective of reconstructing full images at the single-cell level, we demonstrate significant performance improvement over previous attempts on the breast cancer tissue microarray dataset. Our approach offers users access to a more extensive set of biomarkers beyond what has been experimentally measured. It also allows for allocating resources toward exploring novel biomarkers and facilitates greater cell type differentiation and disease characterization. Additionally, it can handle assay failures such as low-quality markers, technical noise, and/or tissue loss in later rounds as well as artificially upsample to include additional panel markers.

1 Introduction

Emerging Multiplexed Tissue Imaging (MTI) platforms such as CO-Detection by indEXing (CODEX) [1], Multiplexed Ion Beam Imaging (MIBI) [2], multiplex immunohistochemistry (mIHC) [3], and cyclic immunofluorescence (CyCIF) [4] produce rich, spatially resolved protein expression information that enables analysis of tissue samples at subcellular resolution. While multiplexed tissue imaging could provide clinically valuable information [5], current MTI approaches are limited by cost, and adoption of these technologies has been limited by prohibitively high material costs and specialized equipment and human expertise required to conduct such assays. Because of these factors, many barriers exist to the wider adoption of MTI workflows into the cancer research community and routine diagnostic tools. Furthermore, the number of markers that can be included in MTI panels is constrained by time both in terms of image acquisition and individual marker selection and validation, and gradual tissue degradation that occurs due to repeated cycles of marker staining and removal [4]. Consequently, the choice of which markers to include in the panel is of utmost importance, in which the goal is to select a set of biomarkers that can be used to identify the widest range of cell states and phenotypes possible.

To assist in this panel selection process, we propose a computational approach to identify markers that can be easily imputed by other co-expressed markers, and impute a surrogate *in silico* high-plex CyCIF from only a few experimentally measured biomarkers. The ability to do so reliably would

Preprint. Under review.

enable substituting easy-to-predict markers with experimental markers that can potentially identify rarer cell types. We propose a deep learning model that can be used to explore the effectiveness of various panel sets of different sizes to reconstruct the full panel in inference, as opposed to retraining a different model for each input panel set. This improvement allows us to search the combinatorial space of marker candidates more efficiently. We show a significant improvement in performance over our prior work [6] on a Breast Cancer Tissue Microarray (TMA) dataset.

1.1 Prior Work

Inspired by the success of masked language modeling models [7] in the natural language processing domain, masked image modeling [8]–[10], has become a popular model pre-training paradigm in computer vision. Masked image modeling can be considered a form of denoising autoencoder [11], which has been used for data restoration and model pre-training. Despite the similarity between masked autoencoding and denoising autoencoding, surprisingly little work has been done to apply masked image modeling to missing data imputation tasks. Herein, we show that the model architecture and masked patch prediction task outlined in [8] can effectively impute CyCIF image channels cropped to the single-cell level, enabling marker stain imputation by way of 'channel in-painting'.

Prior work has been done to identify an optimally reduced panel set in MTI [6], [12], [13]. Our previous work [6] was the first to demonstrate CyCIF panel selection and imputation. We employed a two-step approach by exploring multiple strategies for selecting the reduced panel, then using the reduced panel as input to train a multi-encoder variational autoencoder (ME-VAE) [14] to reconstruct the full panel of 25-plex CyCIF images cropped to the single cell level. Wu *et al.* [12] proposed a three-step method of first using a concrete autoencoder [15] to find a reduced panel of 7 CODEX markers from a full set of 40, then using a convolutional neural network [16] to learn morphological features from the reduced marker set using a multi-scale, single-cell resolution image as input. They then feed those features to a regression model to get predicted mean intensity values for the withheld markers. Sun *et al.* [13] implemented an iterative marker selection procedure, where a U-Net is repeatedly trained to reconstruct patch-level images, which are used to identify the next best marker to add to the reduced panel.

Our approach differs from previous studies[12], [13] including our previous work[6], which utilize a panel selection strategy that is done separately from and prior to the full panel reconstruction process. Additionally, each strategy results in just one reduced panel of a specific size, limiting out-of-the-box use cases for pre-trained models. Here, we employ a similar approach of iterative marker selection, however, we train our model just once and select optimal markers in inference. This means that a single trained model can be used to generate marker expressions from any subset of markers in the training set. Thus, the proposed approach is a more efficient and reliable method for training a versatile model that can be practically useful.

2 A Masked Image Modeling Approach

We observe that the pre-training objective, originally proposed in [8], can be effectively extended to incorporate channel-wise masking without requiring modifications to the original Masked Autoencoder (MAE) architecture. The MAE architecture is composed of two Vision Transformers (ViT) [17], which function as an encoder and decoder respectively as shown in Figure 1. Concretely, if we have a multi-channel image with dimension $C \times H \times W$ and $C = a \times a$ for some integer a, then the image can be resized to $a \cdot H \times a \cdot W$, and with the patch size set to $H \times W$. As an example, if we have 25 CyCIF markers (i.e., $C = 25$), we can generate 5×5 image patches as shown in Figure 1 (bottom).

The original patch-wise masking strategy in [8] is now a channel-wise masking strategy in our setting, with each patch corresponding to a single channel of the CyCIF image. With this minor modification, the MAE training objective now becomes directly applicable to marker imputation in CyCIF and other MTI platforms where the marker expression for each cell is obtained by measuring the mean intensity within the cell boundary of the image channel corresponding to the marker. Therefore, training a model to reconstruct masked image channels results in its capability to infer marker expressions from a reduced subset of markers.

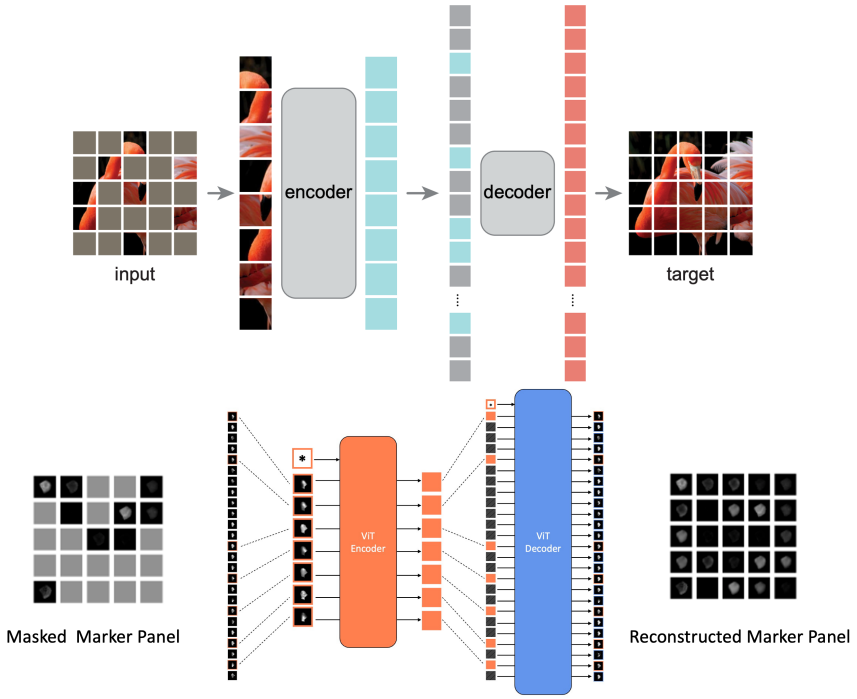

Figure 1: Overview of a masked image modeling approach. (Top) the original Masked Autoencoders (MAE) architecture [8]. (Bottom) we adopt MAE for CyCIF panel reduction and imputation by considering a channel-wise masking strategy with Vision Transformers (ViT) architecture [17].

By implementing a random masking strategy [8], each multi-channel single-cell image in the training set is masked with a different set of channels during each epoch. During inference, we fix the masked channels and observe the effect of different channel selections on the resulting reconstruction. In order to choose an optimal subset of CyCIF makers that results in the best reconstruction of the full marker panel, we seek to utilize a strategy that leverages the power of our pre-trained model. Existing previous works [12]–[14] all operate by applying a panel selection strategy prior to, and independent of full panel reconstruction. This decoupling of reduced panel selection and full panel reconstruction could impede the ability to find an optimal reduced panel as the panel selection strategy is not directly tied to reconstruction. The main advantage of our approach is that it enables the selection of marker panels to be performed during inference, thus avoiding the problem of conflicting objectives between selecting reduced panels and reconstructing the full panel. In addition, this approach allows us to directly address the combinatorial search space of marker sub-panel selection. This improvement is enabled by the MAE training objective, allowing the model to learn general relationships between different subsets of markers at once by random masking at a specific ratio.

Given the aforementioned benefit, our aim is to leverage the generalized model learned by repeatedly probing the model to iteratively identify the marker panels that contribute the most to better reconstructions of the full marker panel.

3 Iterative Panel Selection Algorithm

Given the aforementioned benefit, our aim is to leverage the generalized model learned by repeatedly probing the model to iteratively identify the marker panels that contribute the most to better reconstructions of the full marker panel. Algorithm 1 demonstrates our iterative marker selection approach. We first fix the first marker selected to be a nuclear counterstain (DAPI) due to its importance in

pre-processing steps including nuclei segmentation and image registration of CyCIF:

$$Panel_1 = \{c_{DAPI}\}$$

We then test each candidate marker by including it in the candidate reduced panel and evaluating its influence on full panel reconstruction by measuring the mean Spearman correlation of the predicted marker intensities. The marker that produces the highest correlation between the real and predicted withheld marker intensities is added to the optimal panel set and this process is repeated iteratively until we find the easiest-to-predict marker as follows:

$$Panel_n = Panel_{n-1} \cup \{\arg\max_c(MAE(X, Panel_{n-1} \cup \{c\}; \theta))\}$$

where X represents data, c is a marker not already included in the reduced panel of size $n-1$ and θ refers to the trained model parameters, implying that selection is made in inference. Note that $\arg\max$ refers to the inferred markers' maximum average Spearman correlation.

Our proposed algorithm for iterative panel selection allows us to probe a pre-trained model to discover the optimal order in which markers should be included.

Algorithm 1 Iterative Panel Selection

Input: data X, panel size n, trained model f
Initialize $bestPanel = \{c_1\}$
for $s = 2$ **to** $n - 1$ **do**
  Initialize $bestScore = -\infty$, $bestMarker = \emptyset$
  for $i = 1$ **to** n **do**
    if $c_i \in bestPanel$ **then**
      $continue$
    end if
    $candidatePanel = bestPanel \cup \{c_i\}$
    $candidateScore = f(X, candidatePanel)$
    if $candidateScore > bestScore$ **then**
      $bestScore = candidateScore$
      $bestMarker = c_i$
    end if
  end for
  $bestPanel = bestPanel \cup \{bestMarker\}$
end for

4 Results

4.1 Performance evaluation

In our previous work [6], the optimally reduced panels were found by grouping the markers that maximized the correlation to all the markers withheld from the panel. We show that MAE outperforms our previous approach on the same intensity correlation-based panels as shown in Figure 2 where we compare the previous result [6], the result from MAE approach using the same panel (correlation-based approach in [6]) and the result from MAE approach using iterative panel selection. The full imputations were evaluated by comparing them to the original CyCIF full panels. In [6], a reduced panel can reconstruct relevant unseen information by achieving a mean Spearman correlation of 0.89 for all markers when 18 of 25 markers (28% reduction) are included in the reduced panel. Our proposed approach achieved similar performance with a 64% panel marker reduction (9 of 25 markers).

Figure 3 shows the real and predicted image pairs based on a reduced panel (9 markers). As can be seen qualitatively, the morpho-spatial features of size, shape, marker localization and distribution, and relative intensity are preserved. We also measure the structural similarity index measure (SSIM) quantitatively, which is a widely used measure of image similarity as perceived by the human visual system. The overall quantification shows a mean SSIM of 0.90 in our proposed approach based on a reduced panel (9 markers).

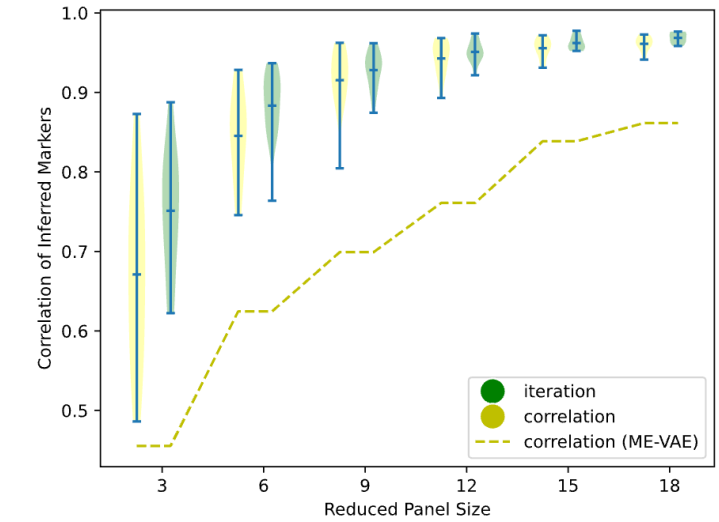

Figure 2: MAE outperforms the previous approach [6] (dashed line) on the same panels (i.e., correlation-based) and further improves performance with the proposed iterative panel selection. Spearman correlation was measured for each stain independently across the multiple reduced panels.

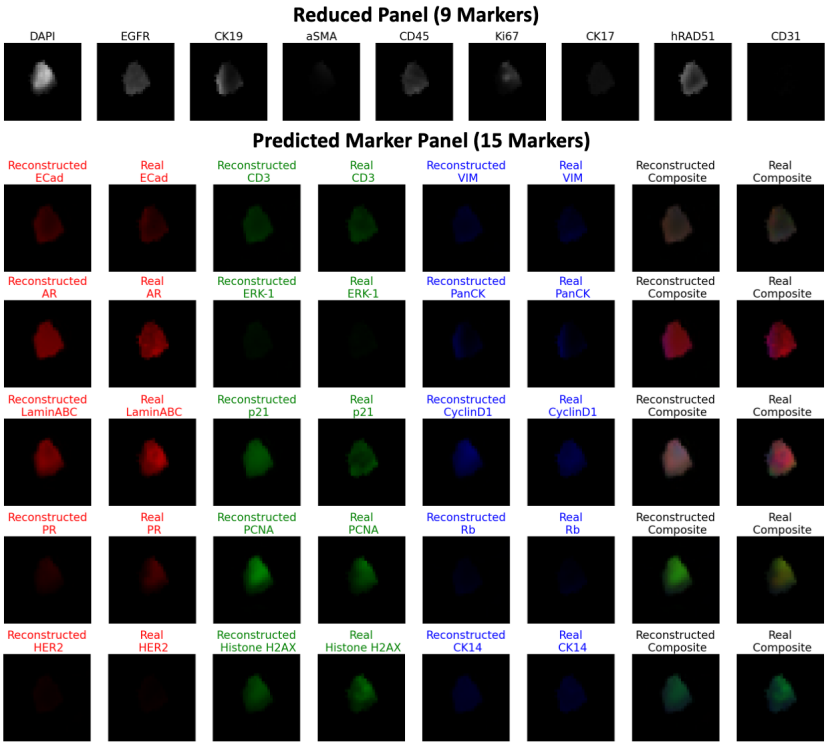

Figure 3: Inferred marker channels retain structural information. (Top) optimally designed reduced panel (9 markers) (Bottom) Predicted marker channel (selected 15 markers instance).

4.2 Iterative marker selection

Iterative marker selection in inference reveals optimally reduced panel sets. Additionally, we find that our iteratively constructed panels further boost performance. In contrast to [6], MAE achieves these results without retraining the model for a specific panel.

Figure 4 shows the iterative panel selection procedure and result. Each pixel represents the Spearman correlation between ground truth and predicted marker intensities. Each column represents the results produced by a different reduced panel size, denoted on the x-axis by the next marker that was chosen in the selection process, for example, the first column represents the prediction results of 24 withheld markers, using a reduced panel of just one marker (DAPI). Rows track the improvement in the prediction of each withheld marker as more markers are included in the reduced panel, for instance, the prediction for PanCK improves significantly in column 2 when ECad is added to the panel. Figure 5 further demonstrates the prediction results for a set of 10,000 randomly selected cells. Each row of plots depicts predictions for witheld markers from a reduced panel of 3, 6, 9, 12, and 15 markers, respectively.

4.3 Masking ratio selection

Although the model was trained using a fixed ratio of masked channels, we find that the model performs well on a range of different ratios in inference. To determine how the masking ratio affects the overall prediction of CyCIF markers, we trained our model with different masking ratios (25%, 50%, and 75%) and assessed the performance of marker prediction for each ratio during inference with varying panel sizes as shown in Figure 6.

Our study shows that a 50% masking ratio yields the best average performance across masking ratios in inference. However, when the masking ratio is high, a 75% masking ratio results in better

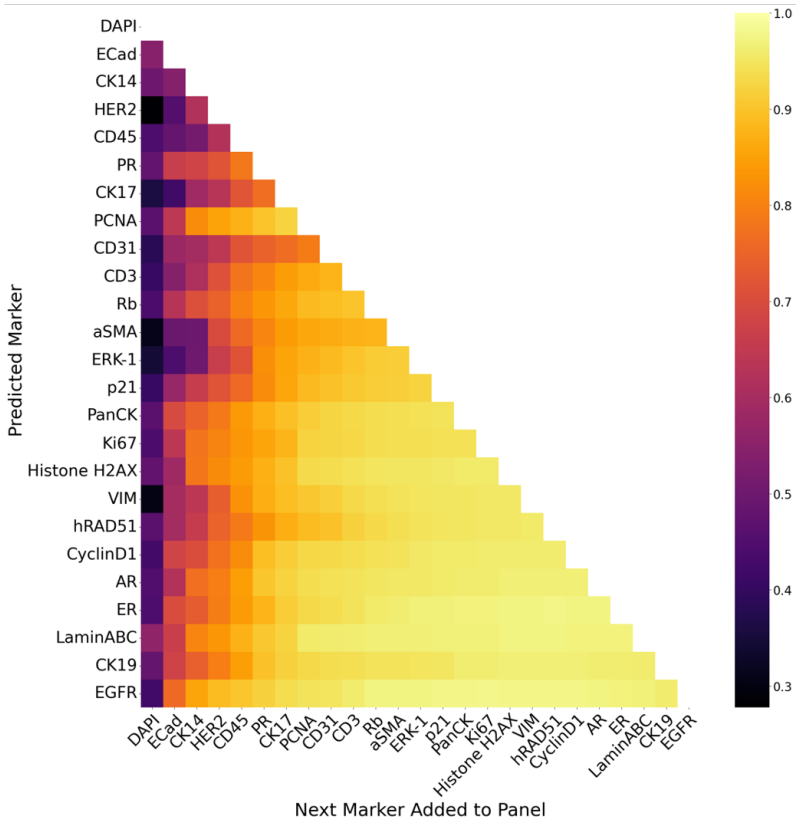

Figure 4: Optimally reduced panels and their inferred marker Spearman correlations.

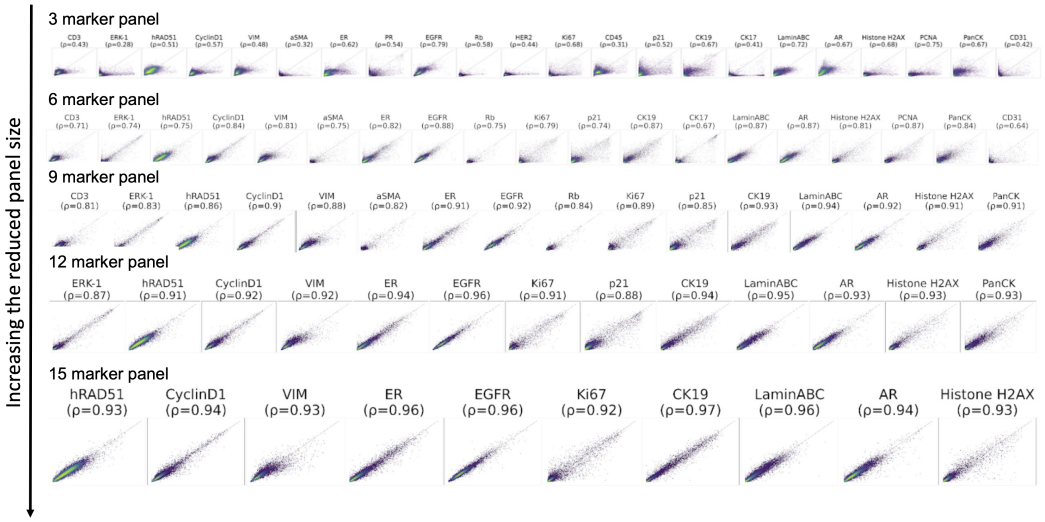

Figure 5: Real versus predicted mean intensity values from a reduced panel of 3, 6, 9, 12, and 15 markers with associated Spearman correlations.

performance, but at the cost of worse performance at lower masking ratios. This aligns with the assumption that the model has not been provided with enough information during training for the 75% ratio. Interestingly, we observed that providing too much information to the encoder during training results in worse model performance during inference, as demonstrated by the lower performance when using a 25% ratio in training. Overall, our finding suggests that the optimal masking ratio during training and inference depends on the specific task and the level of information required by the model.

5 Discussion

Our study highlights the effectiveness of an MAE approach for generating high-plex CyCIF *in silico* from only a few experimental measurements, significantly reducing required biomarkers. With just 9 markers needed instead of the original panel of 25 (64% reduction), our approach offers several benefits. It provides prompt access to a wider range of biomarkers beyond those experimentally measured, manages assay failures, and saves resources for measuring more biologically relevant biomarkers for effective identification of cell types and disease characteristics[12]. Additionally, it can handle situations where assays fail due to factors such as low-quality markers, technical noise, or potential tissue loss in subsequent CyCIF rounds.

Our approach offers several benefits, including wider availability of insights from CyCIF, simplified technical challenges, and improved clinical histopathology workflows. Virtual staining [18], [19] with reduced panels could improve opportunities for transition to the clinic. Moreover, technologies such as Orion from Rarecyte [20] can measure up to 16-18 markers in one cycle using spectral deconvolution, which can be combined with our approach to enabling even greater cell type differentiation, increasing the biomarker measurements from 20 to $32 (\approx 20 \times 1.64)$ or more.

While the method empowers users to access a broader array of biomarkers beyond those experimentally obtained, the success of *in silico* predictions relies on the ability of the model to generalize from the limited experimental data. In our future work, we plan to explore a more diverse training dataset, encompassing whole slide images from different batches, effectively addressing TMA sampling bias and batch effects [21], [22].

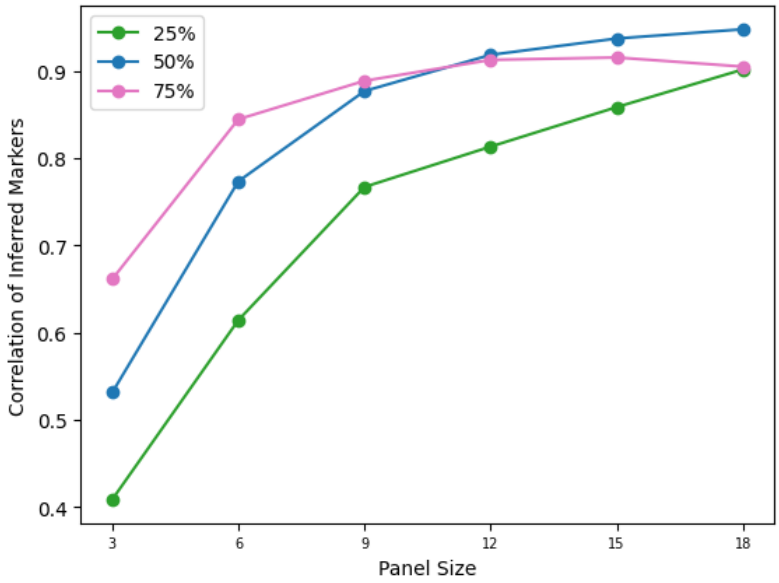

Figure 6: Influence of masking ratio on overall CyCIF marker prediction.

6 Code and Data Availability

All code necessary to reproduce these experiments, including model checkpoints, will be available at `https://github.com/zacsims/IF_panel_reduction`. The Human Tumor Atlas (HTAN) TNP-TMA dataset is available at `https://www.synapse.org/#!Synapse:syn22041595/wiki/603095`.

Acknowledgements

This work was partly supported by the National Cancer Institute – U54CA209988, U2CCA233280, R01 CA253860, and Kuni Foundation Imagination Grants.

References

[1] Y. Goltsev, N. Samusik, J. Kennedy-Darling, *et al.*, "Deep profiling of mouse splenic architecture with codex multiplexed imaging," *Cell*, vol. 174, no. 4, pp. 968–981, 2018.

[2] M. Angelo, S. C. Bendall, R. Finck, *et al.*, "Multiplexed ion beam imaging of human breast tumors," *Nature medicine*, vol. 20, no. 4, pp. 436–442, 2014.

[3] T. Tsujikawa, S. Kumar, R. N. Borkar, *et al.*, "Quantitative multiplex immunohistochemistry reveals myeloid-inflamed tumor-immune complexity associated with poor prognosis," *Cell reports*, vol. 19, no. 1, pp. 203–217, 2017.

[4] J.-R. Lin, B. Izar, S. Wang, *et al.*, "Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-cycif and conventional optical microscopes," *Elife*, vol. 7, 2018.

[5] E. A. Burlingame, J. Eng, G. Thibault, K. Chin, J. W. Gray, and Y. H. Chang, "Toward reproducible, scalable, and robust data analysis across multiplex tissue imaging platforms," *Cell reports methods*, vol. 1, no. 4, p. 100 053, 2021.

[6] L. Ternes, J.-R. Lin, Y.-A. Chen, J. W. Gray, and Y. H. Chang, "Computational multiplex panel reduction to maximize information retention in breast cancer tissue microarrays," *PLoS computational biology*, vol. 18, no. 9, e1010505, 2022.

[7] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[8] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, "Masked autoencoders are scalable vision learners," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 16 000–16 009.

[9] H. Bao, L. Dong, S. Piao, and F. Wei, "Beit: Bert pre-training of image transformers," *arXiv preprint arXiv:2106.08254*, 2021.

[10] M. Chen, A. Radford, R. Child, *et al.*, "Generative pretraining from pixels," in *International conference on machine learning*, PMLR, 2020, pp. 1691–1703.

[11] C. Zhang, C. Zhang, J. Song, J. S. K. Yi, K. Zhang, and I. S. Kweon, "A survey on masked autoencoder for self-supervised learning in vision and beyond," *arXiv preprint arXiv:2208.00173*, 2022.

[12] E. Wu, A. E. Trevino, Z. Wu, *et al.*, "7-up: Generating in silico codex from a small set of immunofluorescence markers," *PNAS nexus*, vol. 2, no. 6, pgad171, 2023.

[13] H. Sun, J. Li, and R. F. Murphy, "Data-driven optimization of biomarker panels in highly multiplexed imaging," *bioRxiv*, pp. 2023–01, 2023.

[14] L. Ternes, M. Dane, S. Gross, *et al.*, "A multi-encoder variational autoencoder controls multiple transformational features in single-cell image analysis," *Communications biology*, vol. 5, no. 1, p. 255, 2022.

[15] A. Abid, M. F. Balin, and J. Zou, "Concrete autoencoders for differentiable feature selection and reconstruction," *arXiv preprint arXiv:1901.09346*, 2019.

[16] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[17] A. Dosovitskiy, L. Beyer, A. Kolesnikov, *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[18] E. A. Burlingame, A. A. Margolin, J. W. Gray, and Y. H. Chang, "Shift: Speedy histopathological-to-immunofluorescent translation of whole slide images using conditional generative adversarial networks," in *Medical Imaging 2018: Digital Pathology*, SPIE, vol. 10581, 2018, pp. 29–35.

[19] E. A. Burlingame, M. McDonnell, G. F. Schau, *et al.*, "Shift: Speedy histological-to-immunofluorescent translation of a tumor signature enabled by deep learning," *Scientific reports*, vol. 10, no. 1, pp. 1–14, 2020.

[20] J.-R. Lin, Y.-A. Chen, D. Campton, *et al.*, "Multi-modal digital pathology for colorectal cancer diagnosis by high-plex immunofluorescence imaging and traditional histology of the same tissue section," *bioRxiv*, pp. 2022–09, 2022.

[21] Y. H. Chang, K. Chin, G. Thibault, J. Eng, E. Burlingame, and J. W. Gray, "Restore: Robust intensity normalization method for multiplexed imaging," *Communications biology*, vol. 3, no. 1, p. 111, 2020.

[22] C. Harris, J. Wrobel, and S. Vandekar, "Mxnorm: An r package to normalize multiplexed imaging data," *Journal of open source software*, vol. 7, no. 71, 2022.