7-UP: generating *in silico* CODEX from a small set of immunofluorescence markers

Eric Wu [*1,2] **Alexandro E. Trevino** [*1] **Zhenqin Wu** [2] **Kyle Swanson** [2] **Honesty J. Kim** [1] **H. Blaize D'Angio** [1] **Ryan Preska** [1] **Gregory W. Charville** [2] **Piero D. Dalerba** [3] **Umamaheswar Duvvuri** [4] **Jelena Levi** [5] **A. Dimitrios Colevas** [2] **Nikita Bedi** [2] **Serena Chang** [2] **John B. Sunwoo** [2] **Aaron T. Mayer** [1] **James Zou** [1,2]

Abstract

Multiplex immunofluorescence (mIF) assays multiple protein biomarkers on a single tissue section. Recently, high-plex CODEX (co-detection by indexing) systems enable simultaneous imaging of 40+ protein biomarkers, unlocking more detailed molecular phenotyping, leading to richer insights into cellular interactions and disease. However, high-plex data can be slower and more costly to collect, limiting its applications, especially in clinical settings. We propose a machine learning framework, *7-UP*, that can computationally generate *in silico* 40-plex CODEX at single-cell resolution from a standard 7-plex mIF panel by leveraging cellular morphology. We demonstrate the usefulness of the imputed biomarkers in accurately classifying cell types and predicting patient survival outcomes. Furthermore, *7-UP*'s imputations generalize well across samples from different clinical sites and cancer types. *7-UP* opens the possibility of *in silico* CODEX, making insights from high-plex mIF more widely available.

1. Introduction

The tissue microenvironment (TME) is a complex milieu comprising many cell types and heterogeneous cell states. Common techniques for understanding the TME like mass spectrometry (Mann et al., 2001) and flow cytometry (Baumgarth & Roederer, 2000) allow for bulk measurements of many cell biomarkers, but discard valuable spatial information in the process. Recently, multiplexed molecular imaging assays have enabled the quantification of cell types and molecules in their native tissue context. Commercial multiplexed immunofluorescence (mIF) systems are increasingly commonplace in clinical diagnostic and prognostic settings (Berry et al., 2021) but are typically limited to quantifying between 1 and 7 biomarkers (Akoya, 2021).

More recently, mIF techniques such as co-detection by indexing (CODEX) (Goltsev et al., 2018) quantify 40 or more markers in situ, allowing a richer and more holistic characterization of the TME and its underlying cell types and disease processes. However, CODEX systems are significantly more costly and time-consuming to run when compared to most low-plex systems, which limits their wider adoption in clinical settings.

To address this limitation, we introduce *7-UP*, a machine learning framework that generates *in silico* high-plex mIF (30+ biomarkers) from only a panel of seven experimentally measured biomarkers. Whereas typical 7-plex measurements can only resolve up to 5-7 distinct cell types (Berry et al., 2021), the imputed biomarkers from *7-UP* enable the identification of up to 16 cell types. Moreover, the imputed biomarker expressions can predict complex clinical outcomes with accuracy comparable to using experimental measurements from CODEX. *7-UP* generalizes to new cancer types and samples that come from different clinical sites than its training data. Our approach highlights a significant opportunity to use machine learning toward inferring high-dimensional molecular features from commonly available low-plex imaging data.

1.1. Related works

Imputation techniques have been applied to missing data in genomics (Liew et al., 2011; Kim et al., 2005; Hastie et al., 2001) and transcriptomics (Zhou et al., 2020; Hou et al., 2020) datasets, as well as in mass spectrometry and shotgun proteomics (Liew et al., 2011; Emerson et al., 2009; Liu & Dongre, 2021) data. Deep learning has been used to extract morphological and spatial features from pathology H&E-stained slides (Zhu et al., 2016; Alom et al., 2019; Lu et al., 2021), and in turn enabled *in silico* IHC staining (He

[*]Equal contribution [1]Enable Medicine, Menlo Park, CA [2]Stanford University, Stanford, CA [3]Department of Pathology and Cell Biology, Columbia University, New York, NY [4]Department of Otolaryngology, University of Pittsburgh, Pittsburgh, PA [5]CellSight Technologies, San Francisco, CA. Correspondence to: A.T.M. <aaron@enablemedicine.com>, A.E.T <alex@enablemedicine.com>, J.Z. <jamesz@stanford.edu>.

The 2022 ICML Workshop on Computational Biology.

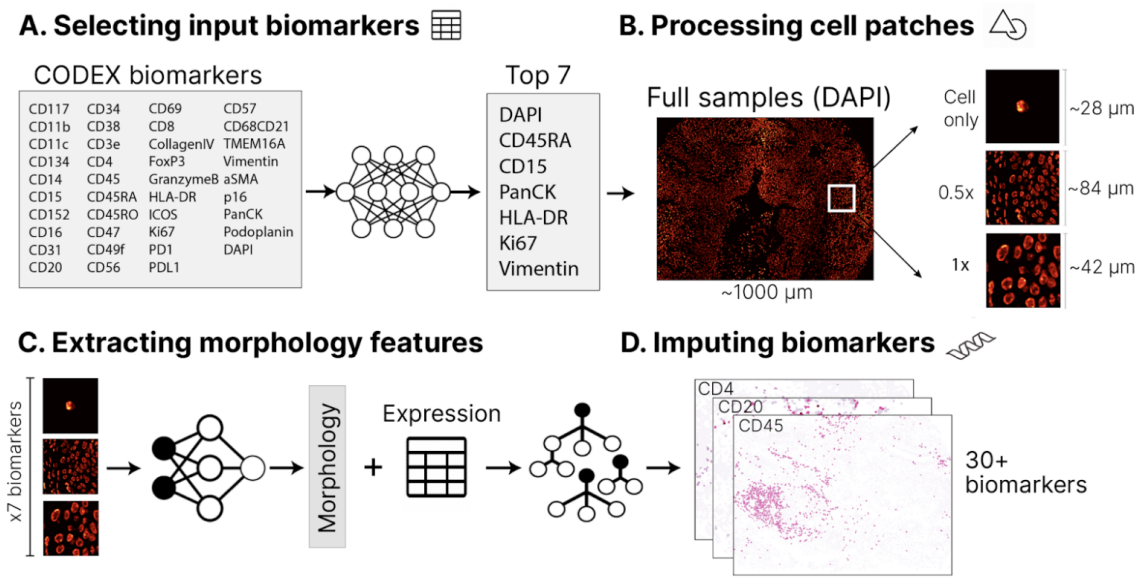

Figure 1. **Overview of the *7-UP* Framework. Panel A**: Seven optimal biomarkers are selected from the full CODEX panel using a deep learning autoencoder model. **Panel B**: From a full sample for each of the seven biomarkers, image patches are extracted for each cell. **Panels C&D**: A deep learning model extracts morphological features from the image patches of each cell, which are combined with the average expressions of the top seven biomarkers to predict the average cell expressions of the remaining CODEX panel biomarkers using a machine learning regression model.

et al., 2022) and spatial transcriptomics (He et al., 2020). More recently, computational methods have been developed for improving cell type classification in CODEX-acquired data (Hickey et al., 2021), and augmenting with spatial information in particular (Zhang et al., 2022). To date, our work is the first to demonstrate the effectiveness of deep-learning-based morphological feature extraction toward multiplex immunofluorescence imputation.

2. *7-UP* summary

The *7-UP* framework consists of the following pipeline: First, we select an optimal panel of 7 biomarkers from the full CODEX biomarker panel. While the choice of which biomarkers to measure in a 7-plex imaging workflow can depend on clinician preference and disease subtype, we use a previously validated approach, Concrete autoencoder (Abid et al., 2019), for automatically selecting informative biomarkers. This approach identified DAPI, CD45RA, CD15, pan-cytokeratin (PanCK), HLA-DR, Ki67, and Vimentin ("Main panel" in Table 1), which we use in our main experiments. Second, we extract cell-level spatial features across each of these seven biomarkers in the CODEX dataset. To do this, we train a convolutional neural network (He et al., 2016) to learn spatial and morphological features from cell image patches generated from the full samples. Third, we combine cell-level spatial features with average biomarker expression values to train a machine learning regression model (Chen & Guestrin, 2016) to impute the expression of the 30+ additional biomarkers.

To validate the veracity of the *7-UP* imputed expressions, we use them to predict cell types. To do so, we use them in place of CODEX-measured expressions in a k-nearest neighbors algorithm used to determine cell type ground truth. In turn, these predicted cell types are used in place of the CODEX-measured ground truth cell types in a graph neural network (Wu et al., 2022) trained to produce sample-level predictions for patient-level survival status, HPV status, and recurrence.

3. Results

3.1. Evaluation across cancer types and clinical sites

Our primary dataset consists of 308 samples from 81 patients with head and neck squamous cell carcinomas at the University of Pittsburgh Medical Center (UPMC-HNC). Two external validation datasets are used: a head and neck squamous cell carcinomas dataset with 38 samples from 11 patients from Stanford University (Stanford-HNC) to demonstrate generalization on the same disease, and a colorectal cancer dataset with 292 samples from 161 patients from Stanford University (Stanford-CRC) to demonstrate generalization to another disease. UPMC-HNC is chosen as the primary training and evaluation dataset as it contains the

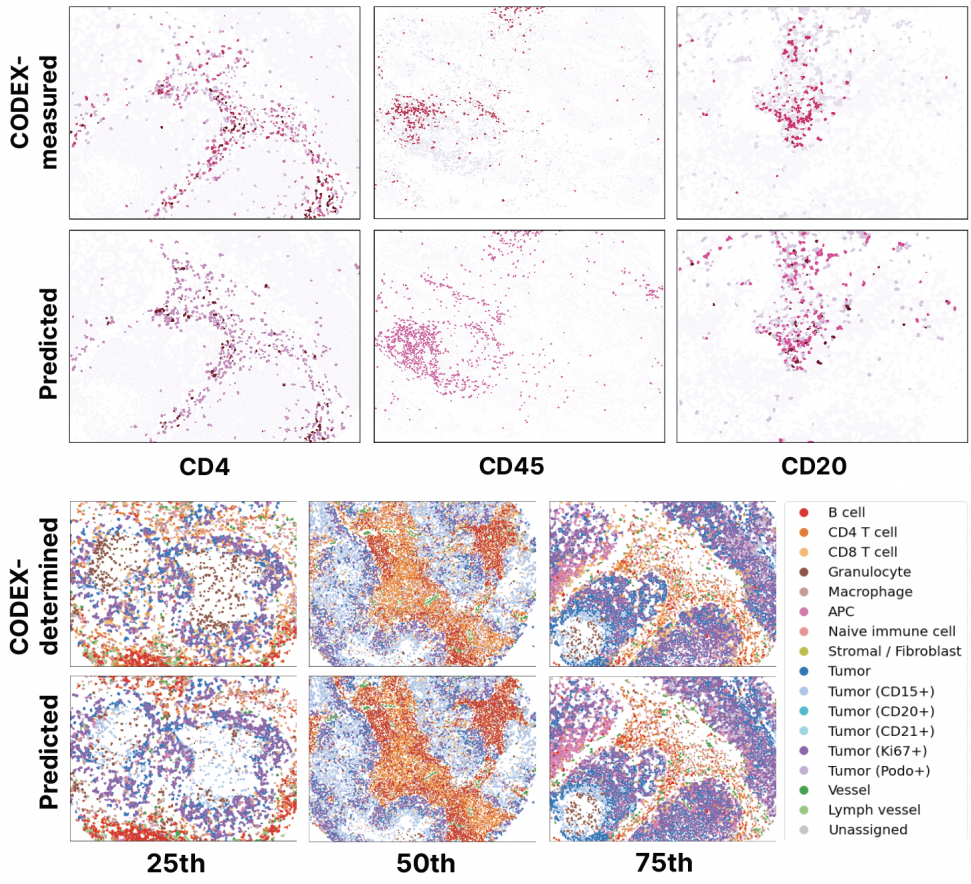

Figure 2. **Top:** Full slide samples of *7-UP* predicted vs. CODEX-measured biomarkers. For each biomarker, the sample with the median patchwise PCC score is shown. **Bottom:** Full slide samples of *7-UP* predicted vs. CODEX-determined cell types. The samples with the 25th, 50th, and 75th percentile predicted patchwise F1 scores are shown.

largest number of samples, coverslips, and total cells. We evaluate our models on held-out coverslips not seen during training to assess model robustness to technical artifacts across coverslips.

3.2. Concordance of biomarker imputations

7-UP achieves an average Pearson correlation coefficient (PCC) of 0.534 across all predicted biomarkers using the main panel in the UPMC-HNC dataset (Table 1). The predictive performance also holds across an alternative input panel (PCC of 0.529), which consists of DAPI, CD4, CD15, PanCK, CD8, Ki67, and Vimentin. Immune-related biomarkers like CD4, CD20, and CD45 are most accurately predicted, with PCCs above 0.70.

3.3. Predicting cell types from imputed biomarkers

We also measure the reliability of the imputed biomarkers by using them for determining cell types since cell type identification is a common task in analyses of CODEX data. Toward this task, *7-UP* achieves a patchwise F1 score of 0.727 (Table 1). The ground truth labels in this analysis are determined with the full CODEX-measured biomarker panel.

3.4. Predicting patient phenotypes from predicted cell types

To validate the reliability of the cell types determined by *7-UP* imputed biomarkers, we use them to predict three patient phenotypic outcomes: HPV infection status, primary outcome (survival), and recurrence of disease. To this end, we use a graph-based deep learning model (Wu et al., 2022) trained using ground truth cell types from the UPMC-HNC dataset to predict these three binary outcomes. We replace the CODEX-measured cell types used to make the baseline prediction with the predicted cell types as input to the model. The results shown in Table 3 demonstrate that the imputed cell types can predict phenotypic outcomes at a level comparable to the ground truth labels.

Table 1. **Performance of *7-UP* on the UPMC-HNC dataset.** Biomarker imputation results are reported using the average patchwise Pearson correlation coefficient (PCC). Cell type predictions are reported using a patchwise weighted F1 score. The first row refers to the imputation regression model trained without including morphological features in the input. The second and third rows refer to the models trained with morphological features. The main and alternative panels are described in the Results section. Numbers in parentheses indicate the 95% bootstrapped confidence intervals.

	PATCHWISE PCC	PATCHWISE F1
UPMC-HNC DATASET	33 BIOMARKERS	16 CELL TYPES
7 BIOMARKERS	0.474 (0.006)	0.667 (0.002)
7 BIOMARKERS + MORPHOLOGY W/ MAIN PANEL	0.534 (0.009)	0.727 (0.002)
7 BIOMARKERS + MORPHOLOGY W/ ALTERNATIVE PANEL	0.529 (0.007)	0.739 (0.002)

Table 2. ***7-UP* generalization**. Imputed biomarker and predicted cell type performance are reported on two external validation datasets (Stanford-CRC and Stanford-HNC). The performance of a model trained on the UPMC-HNC dataset is reported on each validation dataset, along with a reference model that has been trained on the validation dataset. Numbers in parentheses indicate the 95% bootstrapped confidence intervals.

	PATCHWISE PCC	PATCHWISE F1
STANFORD-CRC DATASET	24 BIOMARKERS	16 CELL TYPES
WITH UPMC-HNC MODEL	0.489 (0.024)	0.614 (0.004)
WITH STANFORD-CRC MODEL	0.583 (0.031)	0.605 (0.004)
STANFORD-HNC DATASET	26 BIOMARKERS	18 CELL TYPES
WITH UPMC-HNC MODEL	0.475 (0.005)	0.757 (0.001)
WITH STANFORD-HNC MODEL	0.545 (0.004)	0.773 (0.001)

Table 3. **Phenotype predictions using *7-UP***. Three phenotypic outcomes are predicted using imputed vs CODEX-determined (groundtruth) biomarkers. AUROC scores are reported. Numbers in parentheses indicate the 95% bootstrapped confidence intervals.

AREA UNDER CURVE (AUC)	SURVIVAL STATUS	RECURRENCE	HPV STATUS
CODEX-DETERMINED CELL TYPES	0.889 (0.054)	0.887 (0.108)	0.929 (0.036)
IMPUTED CELL TYPES	0.841 (0.068)	0.894 (0.102)	0.896 (0.051)

3.5. Cross-site and cross-disease generalization

Finally, we evaluate our model on another head and neck cancer dataset (Stanford-HNC) and a colorectal cancer dataset (Stanford-CRC). The biomarker imputation and cell type prediction performances remain stable (Table 2; e.g. for Stanford-CRC: 0.489 vs 0.583 PCC and 0.614 vs 0.605 F1) even when evaluated on a different clinical site and cancer type, indicating that the model's performance is robust when evaluated on unseen data.

4. Discussion

High-plex immunofluorescence techniques like CODEX enable an unprecedented understanding of TME and tissue architecture but have seen limited clinical (diagnostic or prognostic) utility due to their cost and data generation times. On the other hand, standard IF or immunostaining workflows, which image between 1 to 7 biomarkers, are widely used in clinical settings. Our proposed framework aims to unlock the richer TME representations available with CODEX by up-leveling existing 7-plex data through learning biomarker co-expression and morphological patterns.

The ability to determine a subset of biomarkers *in silico* 1) gives users immediate access to a larger set of biomarkers beyond what has been experimentally measured, and 2) frees up resources to measure more novel and biologically relevant biomarkers. Thus, in addition to up-leveling 7-plex systems, *7-UP* can also push CODEX systems beyond 40 biomarker measurements to 60 or more, enabling even greater cell type differentiation and disease characterization.

5. Code and Data Availability

Code will be available at https://gitlab.com/enable-medicine-public/7-up. Datasets are available upon request.

References

Abid, A., Balin, M. F., and Zou, J. Concrete autoencoders for differentiable feature selection and reconstruction. January 2019.

Akoya. Opal 7 solid tumor immunology kit - akoya. https://www.akoyabio.com/phenoimager/assays/opal-7-solid-tumor-immunology-kit/, November 2021. Accessed: 2022-4-14.

Alom, M. Z., Aspiras, T., Taha, T. M., Asari, V. K., Bowen, T. J., Billiter, D., and Arkell, S. Advanced deep convolutional neural network approaches for digital pathology image analysis: a comprehensive evaluation with different use cases. April 2019.

Baumgarth, N. and Roederer, M. A practical approach to multicolor flow cytometry for immunophenotyping. *J. Immunol. Methods*, 243(1-2):77–97, September 2000.

Berry, S., Giraldo, N. A., Green, B. F., ., Szalay, A. S., and Taube, J. M. Analysis of multispectral imaging with the AstroPath platform informs efficacy of PD-1 blockade. *Science*, 372(6547), June 2021.

Chen, T. and Guestrin, C. XGBoost: A scalable tree boosting system. March 2016.

Emerson, J. W., Dolled-Filhart, M., Harris, L., Rimm, D. L., and Tuck, D. P. Quantitative assessment of tissue biomarkers and construction of a model to predict outcome in breast cancer using multiple imputation. *Cancer Inform.*, 7:29–40, 2009.

Goltsev, Y., Samusik, N., Kennedy-Darling, J., Bhate, S., Hale, M., Vazquez, G., Black, S., and Nolan, G. P. Deep profiling of mouse splenic architecture with CODEX multiplexed imaging. *Cell*, 174(4):968–981.e15, August 2018.

Hastie, T., Tibshirani, R., Sherlock, G., Eisen, M., and Botstein, D. Imputing missing data for gene expression arrays. 1, December 2001.

He, B., Bergenstråhle, L., Stenbeck, L., Abid, A., Andersson, A., Borg, Å., Maaskola, J., Lundeberg, J., and Zou, J. Integrating spatial gene expression and breast tumour morphology via deep learning. *Nat Biomed Eng*, 4(8): 827–834, August 2020.

He, B., Bukhari, S., Fox, E., Abid, A., Shen, J., Kawas, C., Corrada, M., Montine, T., and Zou, J. AI-enabled in silico immunohistochemical characterization for alzheimer's disease. *Cell Reports Methods*, pp. 100191, March 2022.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, June 2016.

Hickey, J. W., Tan, Y., Nolan, G. P., and Goltsev, Y. Strategies for accurate cell type identification in CODEX multiplexed imaging data. *Front. Immunol.*, 12:727626, August 2021.

Hou, W., Ji, Z., Ji, H., and Hicks, S. C. A systematic evaluation of single-cell RNA-sequencing imputation methods. *Genome Biol.*, 21(1):218, August 2020.

Kim, H., Golub, G. H., and Park, H. Missing value estimation for DNA microarray gene expression data: local least squares imputation. *Bioinformatics*, 21(2):187–198, January 2005.

Liew, A. W.-C., Law, N.-F., and Yan, H. Missing value imputation for gene expression data: computational techniques to recover missing data from available information. *Brief. Bioinform.*, 12(5):498–513, September 2011.

Liu, M. and Dongre, A. Proper imputation of missing values in proteomics datasets for differential expression analysis. *Brief. Bioinform.*, 22(3), May 2021.

Lu, M. Y., Chen, T. Y., Williamson, D. F. K., Zhao, M., Shady, M., Lipkova, J., and Mahmood, F. AI-based pathology predicts origins for cancers of unknown primary. *Nature*, 594(7861):106–110, June 2021.

Mann, M., Hendrickson, R. C., and Pandey, A. Analysis of proteins and proteomes by mass spectrometry. *Annu. Rev. Biochem.*, 70:437–473, 2001.

Wu, Z., Trevino, A. E., Wu, E., Swanson, K., Kim, H. J., Blaize D'Angio, H., Preska, R., Charville, G. W., Dalerba, P. D., Egloff, A. M., Uppaluri, R., Duvvuri, U., Mayer, A. T., and Zou, J. SPACE-GM: geometric deep learning of disease-associated microenvironments from multiplex spatial protein profiles. May 2022.

Zhang, W., Li, I., Reticker-Flynn, N. E., ., Sunwoo, J. B., Nolan, G. P., Engleman, E. G., and Plevritis, S. K. Identification of cell types in multiplexed in situ images by combining protein expression and spatial information using CELESTA reveals novel spatial biology. February 2022.

Zhou, Z., Ye, C., Wang, J., and Zhang, N. R. Surface protein imputation from single cell transcriptomes by deep neural networks. *Nat. Commun.*, 11(1):651, January 2020.

Zhu, X., Yao, J., and Huang, J. Deep convolutional neural network for survival analysis with pathological images. In *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 544–547, December 2016.