Automated Single-Cell Quantification of Breast Cancer Marker Evolution from Serial Immunofluorescence Assays

Approved Category

Computer Science and Systems Software

Rikhil Rao, Grade: 11

Abstract

Despite advances in breast cancer diagnoses and treatment, it continues to be the second leading cause of cancer death in women. Recent interest in managing and prognosticating breast cancer has been centered around examining single-cell and spatial expression patterns on molecular imaging, such as immunofluorescence and immunohistochemistry. However, current techniques for this single-cell quantification are disjointed and steps are not formatted to be compatible.

The purpose of this project is to develop a simple, easy to use, end-to-end pipeline to quantify how much a patient's cancer changes at different timepoints in their treatment. To accomplish this, serial cyclic immunofluorescence images acquired over the course of a patient's breast cancer treatment were gathered from the Human Tumor Atlas Network (HTAN). Two images of different samples were chosen from time point 0 and one image from time point 1 was chosen. Each image consisted of 25 unique channels, including DAPI, HER2, and ER. These images were used to train a patient-specific Variational Auto-Encoder (VAE), which was then used to quantify how much the cancer changed. In order to train the model, individual cells were segmented using an open-source tool called MCMICRO, resulting in around 14,000 cells in each image.

By running these analyses, the tumor phenotype of the patient (e.g., HER2 negative, estrogen receptor [ER] positive) was able to be automatically determined. In addition to this the MCMICRO quantification outputs show that the marker expression changed at a later time point as the patient underwent hormone therapy for their breast cancer. For example, the ER expression decreased 2-fold and the Ki67 expression increased 12-fold. Additionally, the VAE latent space plots show significantly different distributions at different time points, implying that the network was able to encode the baseline single-cell expression patterns.

This research shows that automated single-cell quantification of marker expression in breast cancer can be important to track tumor evolution over time, which can be quantified and visualized using a VAE.

Background

Cancer has become one of the main threats to human health and life. Among all types of cancer, breast cancer is very common among women and diverse in character. About 70 percent of all breast cancers in women have overexpression of receptors (proteins) that attach to estrogen or progesterone. Breast cancers without ER (estrogen receptor) and PR (progesterone receptor) are known as HR (hormone receptor)-negative. Estrogen receptor and progesterone receptor tests are used to help guide breast cancer treatment. Cancer cells and cells within the tumor microenvironment together determine disease progression and response to treatment.

Immunofluorescence is a powerful approach for getting more information about cellular structure. Expressions and locations of a multitude of specific proteins can be assessed, making this process best suited for scientists to solve many cell biological questions. The extent of cellular heterogeneity in breast cancer has been hypothesized to impact diagnosis, response to therapy, and long-term outcome.

Determining tumor heterogeneity and its impact on drug response is essential in the development of personalized therapies. Scientists have recognized that images of cells contain more information than what can be extracted by the human eye. Computer aided image analysis has the potential to make complex information more accessible in diagnostic practice. Artificial Intelligence based computational pathology has shown increased diagnostic accuracy and reduced turnaround times.

Problem Statement

Breast cancer is the second leading cause of cancer death in women and is responsible for 40,000 deaths per year in the United States. Diagnosis and prognostication of breast cancer is done through evaluation of a tumor biopsy using immunofluorescence. Recent advances in analyzing multiplex immunofluorescence at the single cell level have shown cell-specific spatial differences of various tumor marker expressions within the same biopsy, which may be indicative of long-term outcomes. However, current methods of clinical analyses are not able to quickly quantify markers at the single-cell level. In this study, the MCMICRO tool is applied to breast cancer molecular assays in order to automatically quantify the tumor phenotype and quantify single-cell changes in expression over time using a Variational Autoencoder.

Hypothesis

I hypothesize that an automated cell segmentation and quantification algorithm can accurately quantify breast cancer marker phenotypes in single cells and can be applied to track cell expression changes over time.

Materials

Software:

- **Visual Studio** – IDE and code editor for software developers
- **Google Collab** – Allows anybody to write and execute arbitrary python code through the browser
- **Python 3.9** – A high level, general-purpose programming language
- **Keras** – High level API of Tensorflow
- **MCMICRO** – Open-source software that processes whole slide microscopy data into cohesive images that can easily be visualized and quantified as single cell data
- **Google Cloud CLI** – Set of tools to create and manage Google Cloud resources
- **Breast Cancer database** – the Human Tumor Atlas Network (HTAN) database from which imaging data was downloaded

Procedure

To create and train AI model, the following steps were performed.

Downloading the Dataset:

Data was downloaded from the Human Tumor Atlas Network (HTAN) website. The files were downloaded via the Google cloud command line interface (gCloud CLI). All the files were downloaded in dicom format.

Prepare Dataset for MCMICRO:

To run the data through the MCMICRO pipeline, a csv file that contains information on the marker associated with each channel in the dicom image was made. All the dicom files that were downloaded were parsed as they each represented a single channel of the whole image. This information was used to generate a csv file.

Run MCMICRO Pipeline on Dataset:

Using nextcloud, the MCMICRO pipeline was run on the dataset with the generated csv file.

Create training data for VAE:

To create training data for the VAE, the segmentation of the dataset that was performed by the MCMICRO pipeline was used to generate small images (24 by 24 pixels) for each of the cells that were found. The data was then saved into a numpy array.

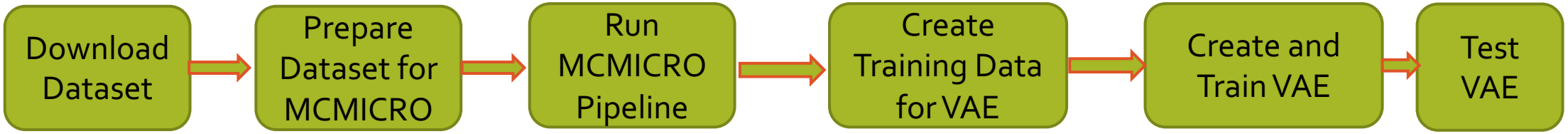
Procedure (Continued)

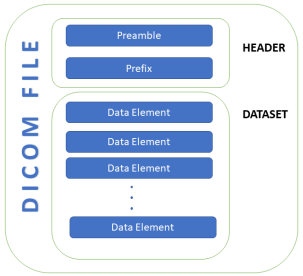
Create and Train VAE:

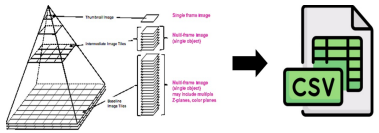
The Variation Auto Encoder was created using keras, a neural network library. The model that was created has 4 2D convolutional layers, 2 dense layers, 1 flatten layer, and 1 2D Convolutional upsampling layer. To train the model, saved data was loaded in the form of a numpy array and then partitioned it so that 75% of the data stored in the file was used for training while the other 25% is used for testing.

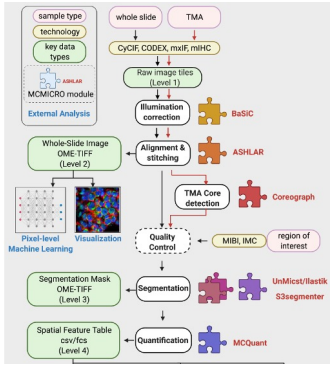
Test VAE:

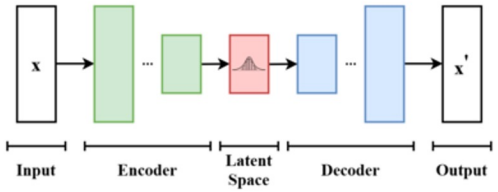
The VAE used data generated from two different timepoints of the biopsy so that the latent vectors of each run could be tested to quantify how much the cancer changed.











Results

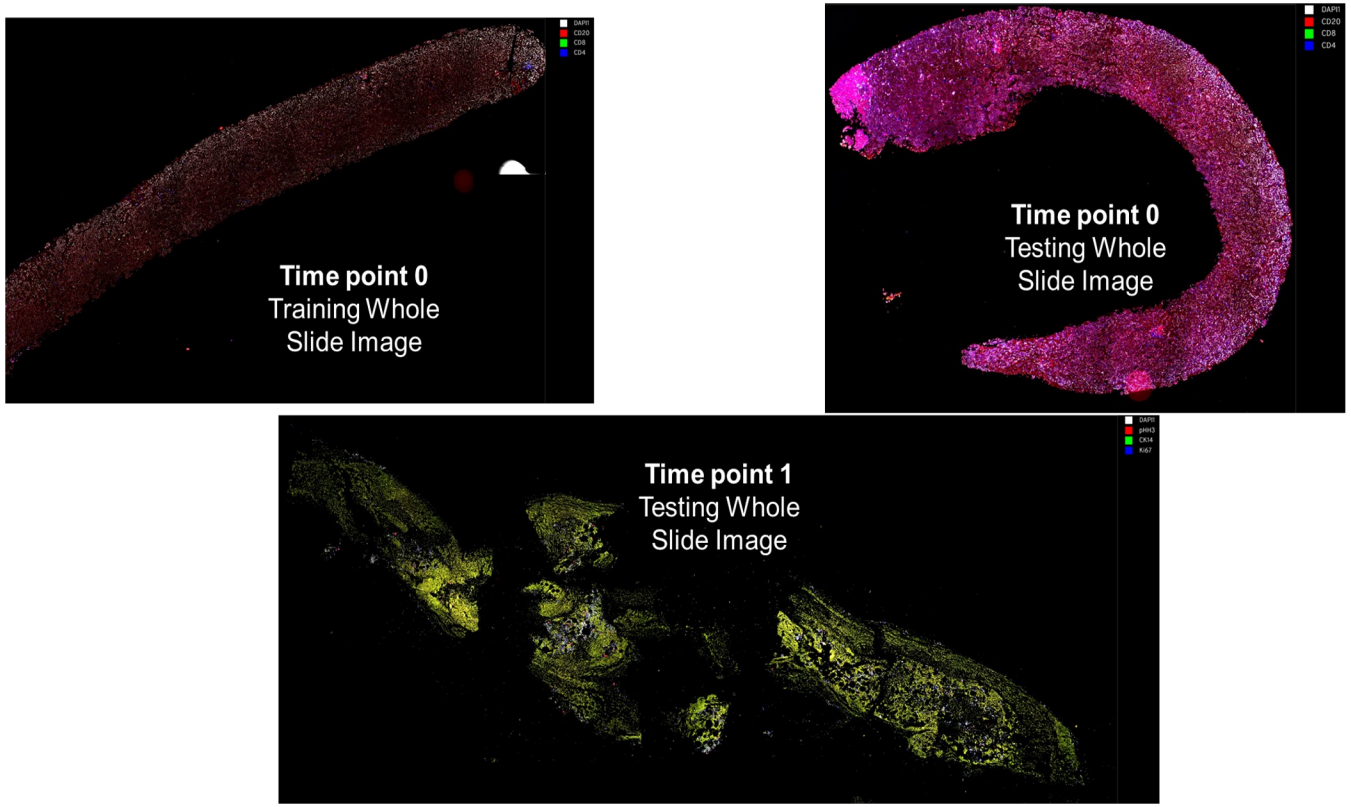
Data

The result is a pipeline that takes in a set of molecular imaging files (often 10-100 GB in size) and outputs the single cell quantification for each channel of the imaging.

In the training dataset, the pipeline took an input of 30 channels (DAPI, HER, ER, PCNA, Ki67, etc.) and accurately segmented ~ 16,000 cells from which cell-specific marker expression quantifications were obtained.

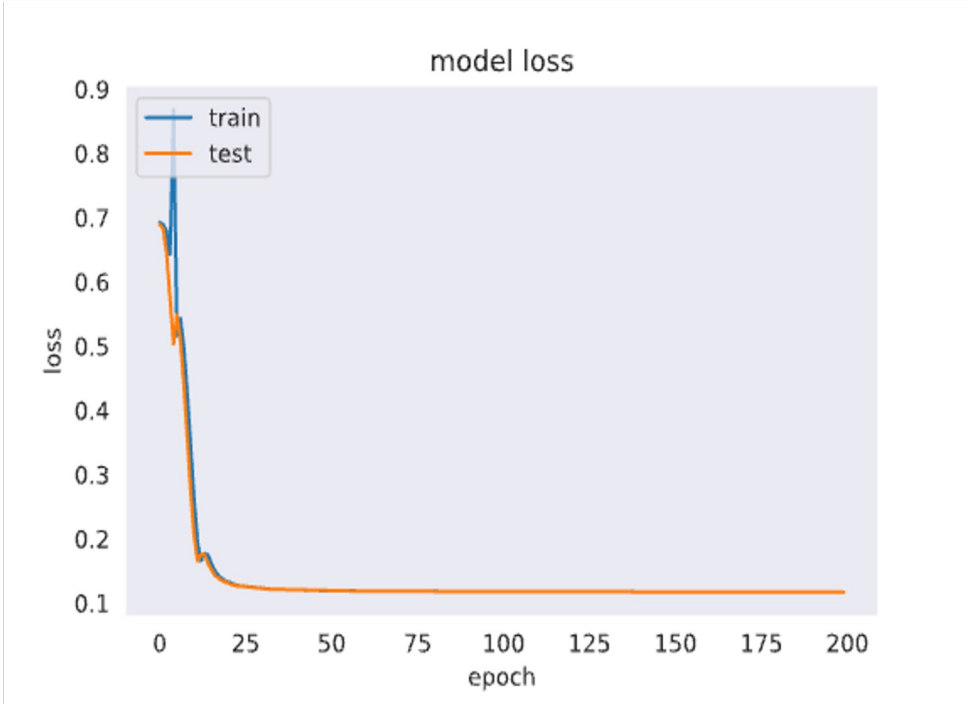
Execution of the pipeline on the testing datasets at the same and later time points resulted in ~21,000 and ~4000 identified and quantified cells, respectively.

Whole Slide Images Used in Study



Results (Continued)

Model Accuracy



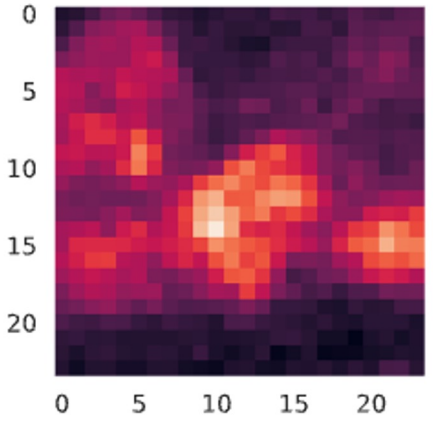
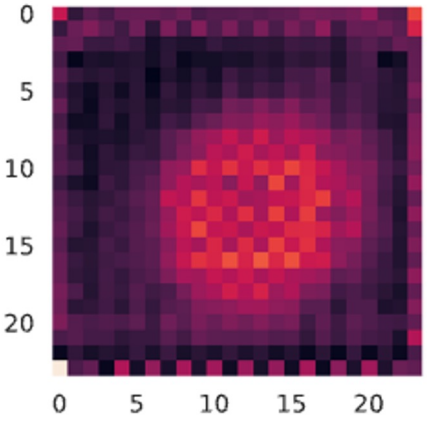


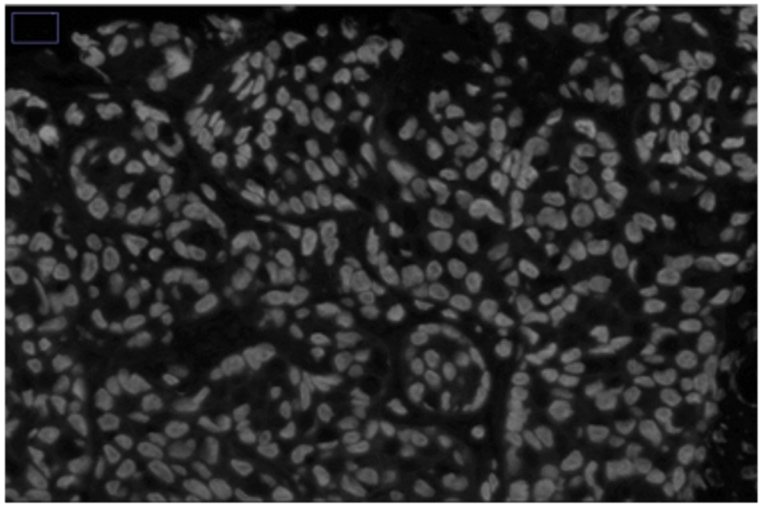
Image fed into VAE



Rebuilt image by VAE

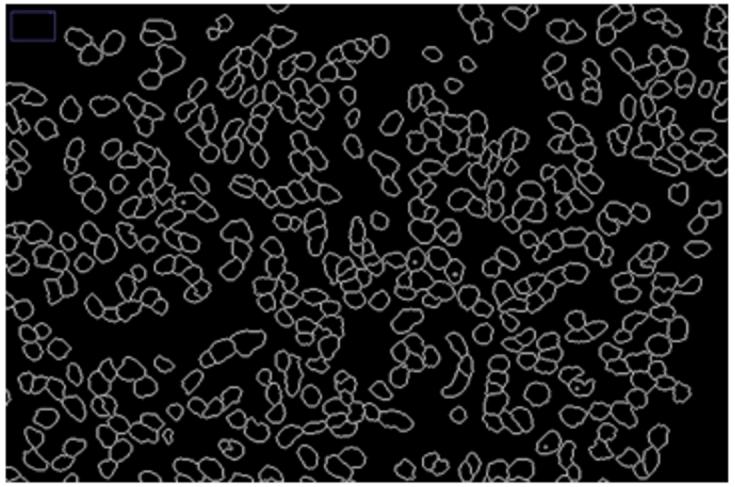
Results (Continued)

MCMICRO Outputs

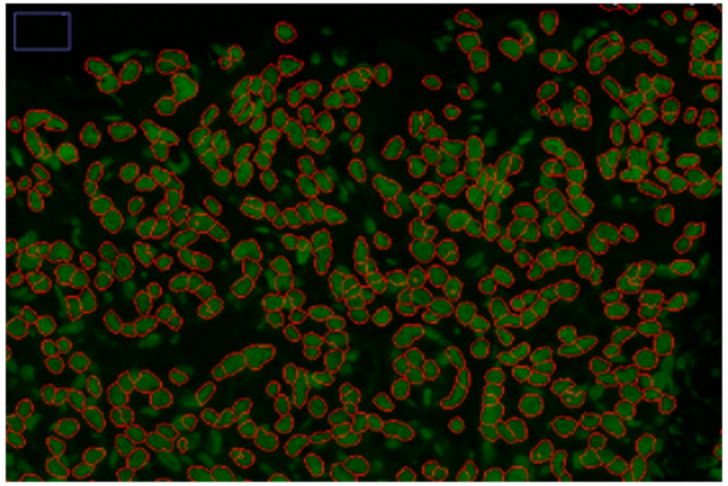


Example ROI for DAPI Channel

MCMICRO UNMICST Algorithm



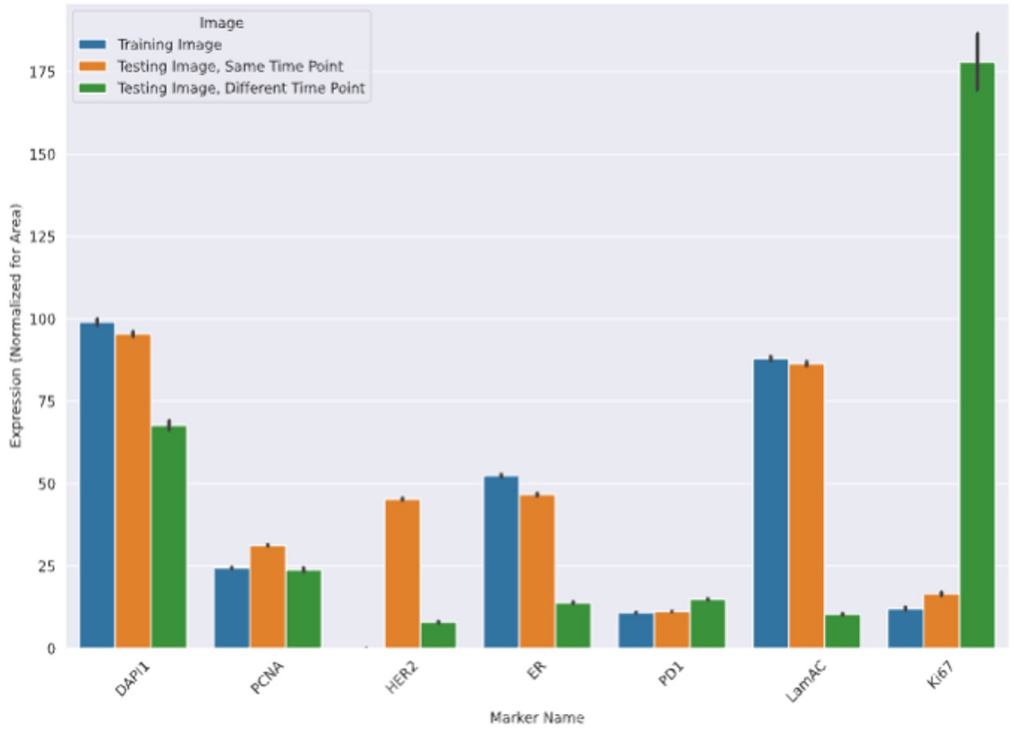
Cell Segmentations



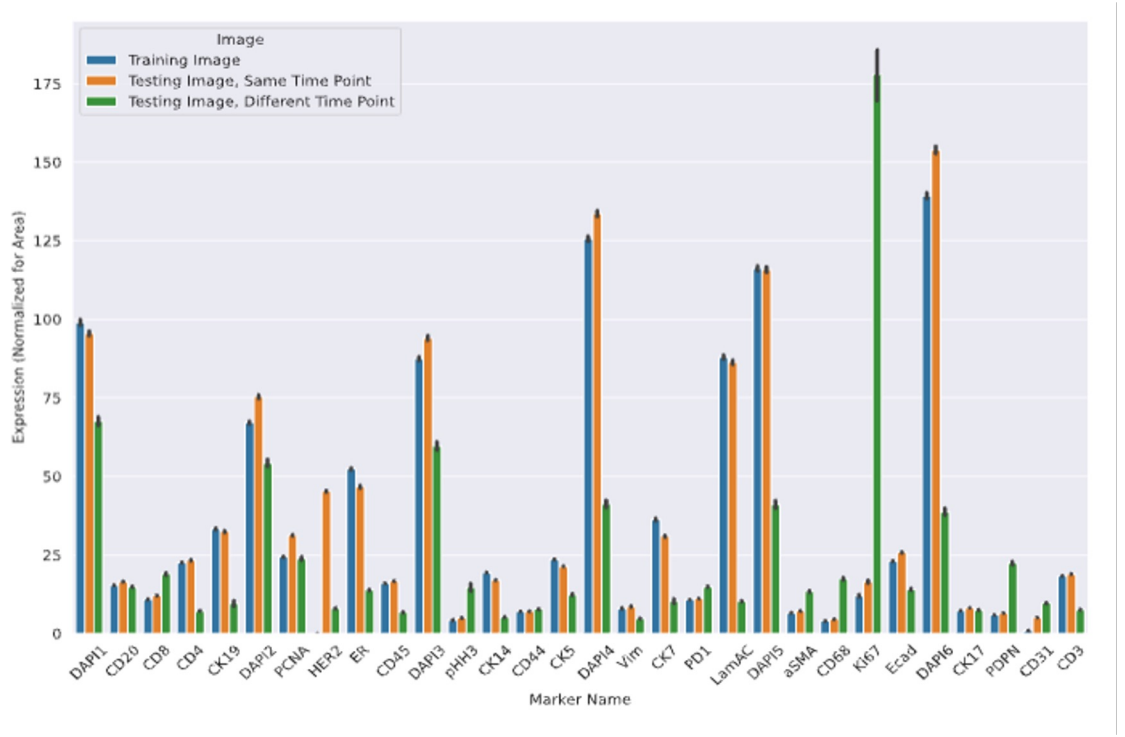
Overlay of Cell Segmentations and DAPI Channel

Results (Continued)

Quantification



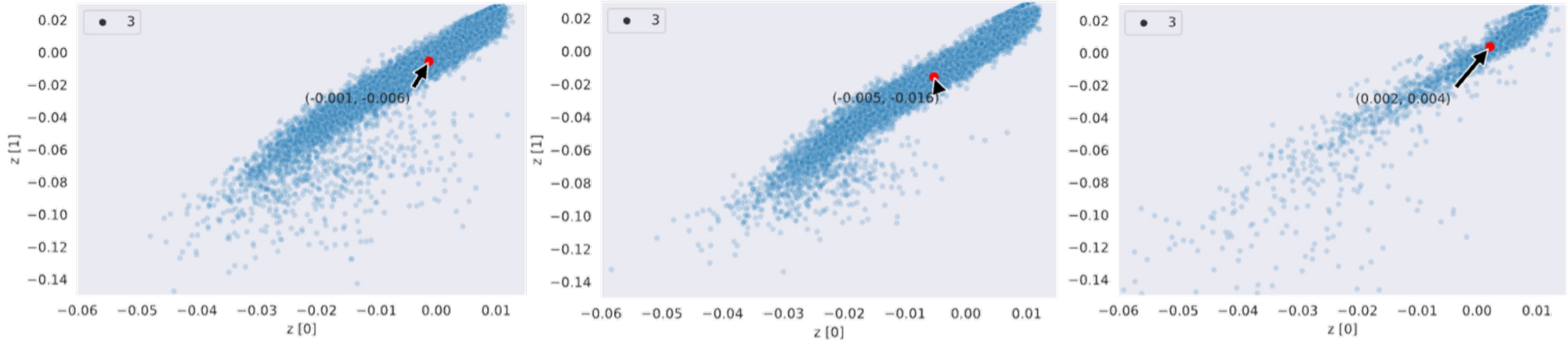
Important Markers



All Markers

Results (Continued)

Model Predictions



The larger KL Divergences for the comparisons *between* timepoints than those *within* a timepoint suggests that the VAE is learning the patient's baseline tumor phenotype.

Conclusion

This project applied a deep learning-based segmentation module to quantify tumor marker expression in breast cancer molecular imaging over time as a patient underwent breast cancer treatment. A VAE trained on baseline data showed that it was able to track and quantify the differences in single-cell expression over time. Future research on this topic should focus on validating this pipeline in multiple other subjects who have serial biopsies and better understanding the implications in tumor progression from the latent space changes.

References

- Characterizing advanced breast cancer heterogeneity and treatment resistance through serial biopsies and comprehensive analytics https://www.nature.com/articles/s41698-021-00165-4
- Impact of mutational profiles on response of primary oestrogen receptor-positive breast cancers to estrogen deprivation: https://www.nature.com/articles/ncomms13294
- Quantitative single-cell analysis of immunofluorescence protein multiplex images illustrates biomarker spatial heterogeneity within breast cancer subtypes: https://breast-cancer-research.biomedcentral.com/articles/10.1186/s13058-021-01475
- Biological and Medical Importance of Cellular Heterogeneity Deciphered by Single-Cell RNA Sequencing – Rishikesh Kumar Gupta and Jacek Kuznicki https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7463515
- Single-cell profiling of tumor heterogeneity and the microenvironment in advanced non-small cell lung cancer - https://www.nature.com/articles/s41467-021-22801-0
- Cancer Ecosystems - https://www.frontiersin.org/articles/10.3389/fonc.2019.00718/full
- Breast Cancer Hormone Receptor Status - https://www.cancer.org/cancer/breast- cancer/understanding-a-breast-cancer-diagnosis/breast-cancer-hormone-receptor-status.html
- Estrogen Receptor, Progesterone Receptor tests - https://medlineplus.gov/lab-tests/estrogen- receptor-progesterone-receptor-tests/
- Artificial Intelligence: How is It Changing Medical Sciences and Its Future? - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC76408
- Precision immunoprofiling by image analysis and artificial intelligence. https://link.springer.com/article/10.1007/s00428-018-2485-z
- Image file https://www.vladsiv.com/dicom-file-format-basics/