

## Background

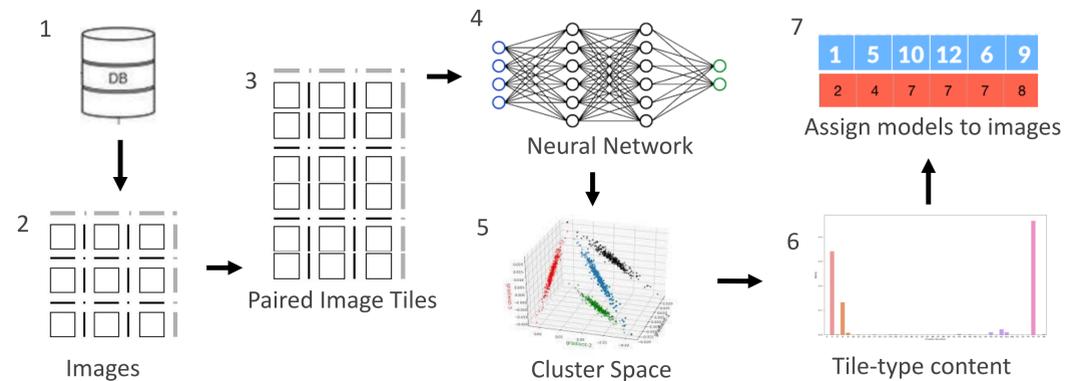
Understanding the tumor composition and microenvironment provides insight into the efficacy and viability of immunotherapy treatments. Obtaining cell-level co-expression data (biomarker patterns expressed on cells indicating their phenotype) provides insight into where and how various cell types are distributed throughout the tumor. The MultiOmyx technology platform provides a detailed quantitative output including co-expression data through a multiplexed fluorescence microscopy approach, followed by imaging analysis using a proprietary analytics pipeline.

The MultiOmyx analytics pipeline involves training a neural network for every biomarker of interest. Each biomarker's neural network classifies all cells as positive or negative for the marker. A major time-sink when constructing these neural network models is manually collecting and annotating training data. Moreover, new models must be built for the same biomarkers across different studies due to variability between samples. However, previously-built models can be applied on new images provided that the training image data used for the model and the new images follow the same morphology, background, and intensity variability. This eliminates the bottleneck step of training a new model. In order to find ideal existing models for a new batch of samples, Invariant Information Clustering can be used to identify similarity of the new batch to previous images.

Invariant Information Clustering (IIC) is used to group historical images by visual appearance. Image data is collected from our internal database, which stores all images from past studies. These images are split into image tiles and are then clustered with IIC. IIC works by generating image pairs that consist of the original image and an augmented version. It then captures the mutual information between the image and its transformed version using a neural network with a Resnet-34 architecture. This results in the extraction of the meaningful parts out of the original image while discarding the instance-specific details. Augmentations including rotations, flipping, cropping, and random noise are used to guide the network to learn proper features and be invariant to others. Visually similar tiles form distinct and tightly-bound cluster groups, each of which has a model assigned to it. When running inference, new image tiles are passed through the network and are assigned to the closest cluster. Cells from each tile are then classified using the model which corresponds to that tile's cluster. If a new tile falls far from an existing cluster, cells from that tile are processed with a new, manually-annotated model.

## Workflow Steps

Our IIC-based framework consists of 7 major steps:

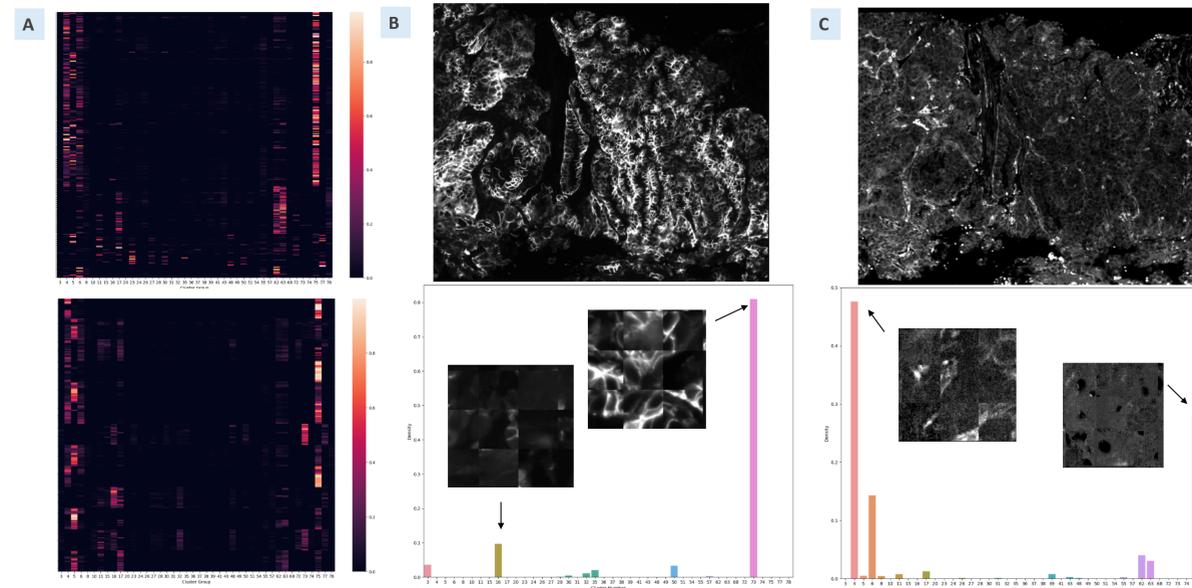


**Figure 1: IIC Workflow:**

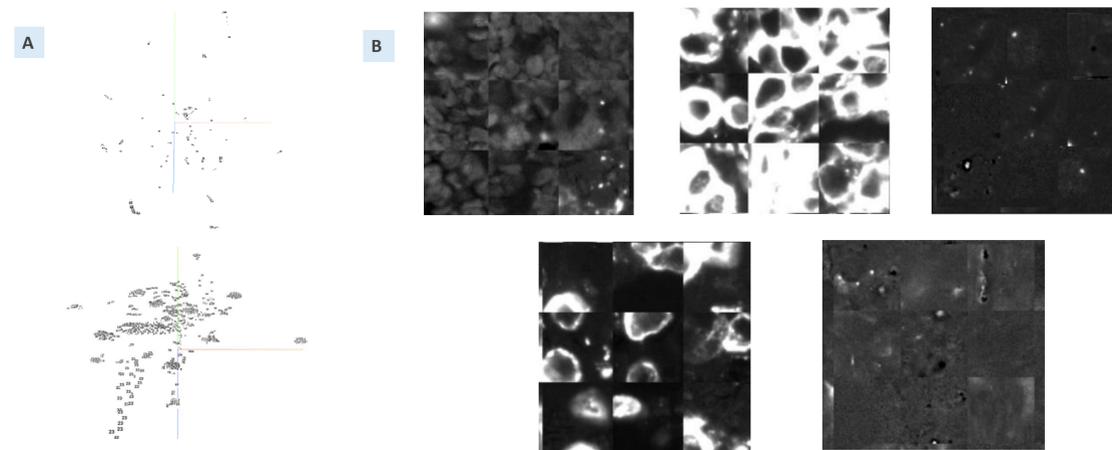
1. Generate large representative sample of images from our MultiOmyx database.
2. Split images into tiles so that they will fit into GPU memory during training.
3. Create image pairs of each original tile and a transformed version for increased robustness and better clusters.
4. Maximize mutual information between each image pair by running images through a Resnet-34 neural network.
5. Perform inference and generate a low-dimensional latent space representation of original input-ROI space where tiles with similar morphology fall into the same clusters.
6. Piece tiles back together and assign each ROI image a vector of length C, the number of clusters. The unique values of the vector correspond to each cluster group and what percent of the ROI's tiles were in that group.
7. Cluster vector representations using k-means clustering to find the most suitable model for each full image.

## Cluster Analysis

To evaluate the clustering quality, we can perform visual inspection of the most heavily populated clusters and look for observable morphological pattern, and observe the clustering quality with the aid of various data visualization techniques.



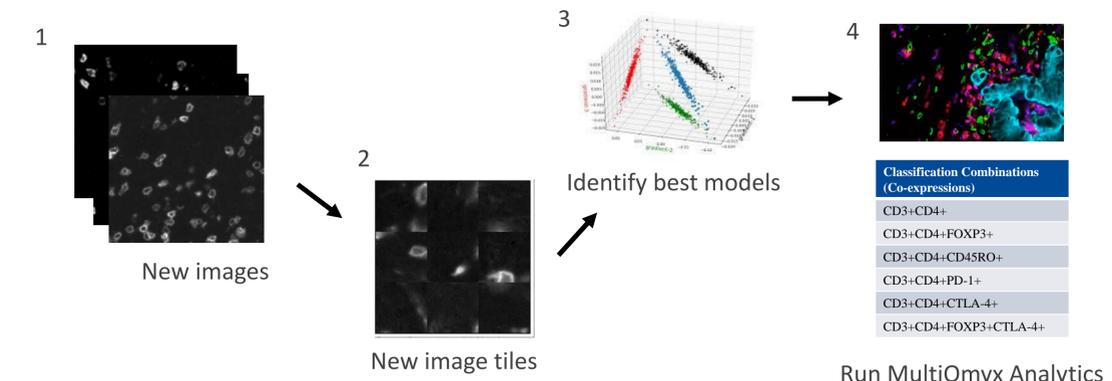
**Figure 2: A.** Heat maps for IIC cluster prevalence across different slides and ROIs for the same biomarker, sorted by slides, then ROIs along the y-axis (top), and across different biomarkers for the same sample, sorted by biomarkers, then ROIs along the y-axis (bottom) **B/C.** CD56 (B) / CD4 (C) biomarker image for a single ROI with IIC cluster histograms and representative tiles from the two most prevalent IIC clusters in each.



**Figure 3: A:** UMAP (top) and t-SNE (bottom) visualizations of clusters from example tiles. **B.** Randomly sampled tile images (of the training set) from five IIC clusters.

## Incorporation with MultiOmyx Workflow

The IIC-based workflow is used to retrieve an optimal model from the pretrained clustering space, which is then used on new images to generate MultiOmyx Analytics outputs.



**Figure 4: MultiOmyx Workflow:**

1. Acquire new images.
2. Tile new images.
3. Identify best models using IIC workflow.
4. Process images using MultiOmyx Analytics workflow with best models.

## Summary

NeoGenomics has developed an unsupervised deep learning pipeline with which to identify optimal historical models and perform the MultiOmyx Analytics pipeline on new samples. As a result, model training and data annotation tailored to new images is not needed. This method eliminates the time used to build new models (except for extreme outlier cases) and expedites the analytics process, thus reducing turnaround time and allowing for a greater volume of samples to be processed. Moreover, this is an iterative process: As more sets of patient samples come in, they can be used to refine the cluster space and improve models in the future.

## References

1. Gerdes MJ, Sevinsky CJ, Sood A, et al. "Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue." *Proc Natl Acad Sci U S A.* 2013 ;110(29):11982-11987. doi:10.1073/pnas.1300136110
2. Ji, Xu, Jo Henriques, and Andrea Vedaldi Kaiming, et al. "Invariant Information Clustering for Unsupervised Image Classification and Segmentation." *Proceedings of the IEEE international conference on computer vision.* (2019): 9865-9874.
3. LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." *nature* 521.7553 (2015): 436-444.
4. He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep Residual Learning for Image Recognition." arXiv: 1512.03385, 2015.