Scalable Informatics Tools for Investigating Intra-Tumor Heterogeneity in Breast Cancer

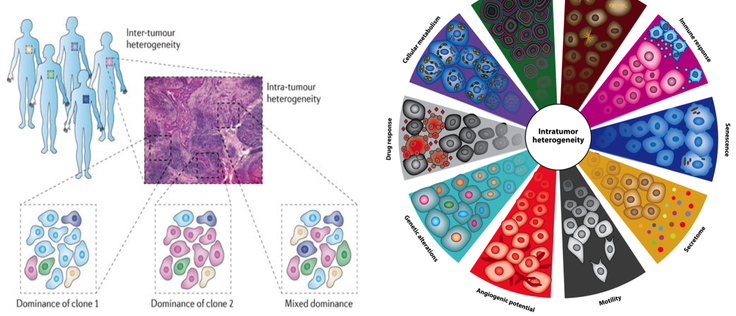
A.J. Andonian[1,2]; D.M. Spagnolo[3,4,5]; L. Nguyen[3,4,5]; D.L. Taylor[4,5]; S.C. Chennubhotla[4,5]

[1]TECBio REU @ Pitt [2]Bates College, Lewiston, Maine [3]Joint Carnegie Mellon University - University of Pittsburgh Ph.D. Program in Computational Biology
[4]Drug Discovery Institute, University of Pittsburgh [5]Department of Computational and Systems Biology, University of Pittsburgh

1. Abstract

Intra-tumor heterogeneity, a prominent feature of many malignancies, may have diagnostic and prognostic value. Thus, it is essential to develop and test spatial intra-tumor heterogeneity metrics that correlate with various clinical outcomes. In the present study, we propose a high throughput pipeline that will form the basis for a set of open-source informatics tools for integrating and analyzing data obtained from a variety of imaging modalities. In particular, we employ a deep learning approach to identify discriminative cellular distributions or “signatures” in multiplexed immunofluorescence images that can be used to characterize breast cancer subtypes. Lastly, our distributed implementation currently targets Apache Spark, a powerful cluster computing framework, and should begin to mitigate the difficulties associated with quantitative big imaging.

2. Intra-tumor Heterogeneity

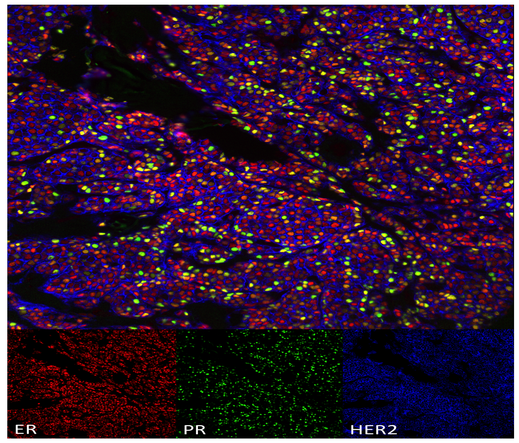


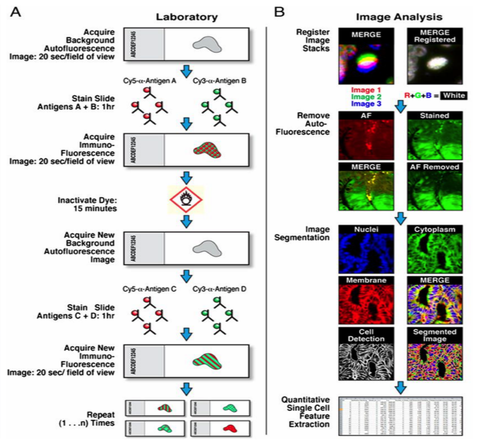
Tumors represent a complex dynamic ecosystem where heterogeneity acts as the substrate for tumor evolution and as one of the main drivers of disease progression and resistance to therapy.

3. Data

- Tissue microarray of 99 samples consisting of triplicate, 1mm diameter cores from 24 invasive breast tumor tissues
- Immunohistochemical expression levels for estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor 2 (HER2) revealed four cohorts:
  - 3 cases of ER (+) invasive ductal carcinoma (IDC)
  - 5 cases of ER (+) invasive lobular carcinoma (ILC)
  - 8 cases of ER (-) IDC
  - 8 cases of HER2 (+) IDC

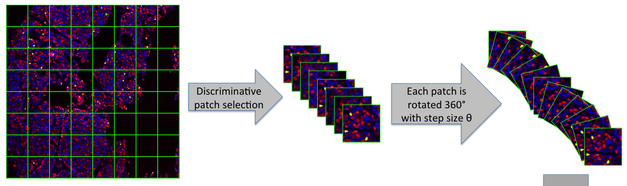
4. Multiplexed Immunofluorescent Imaging





5. Patch-Based Convolutional Neural Network

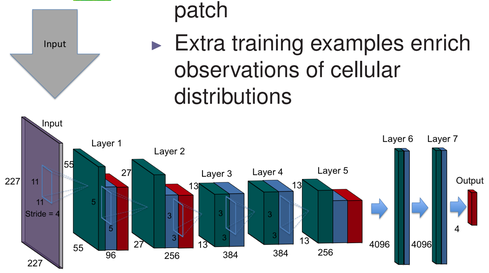
5.1 Rotation-Based Patch Augmentation



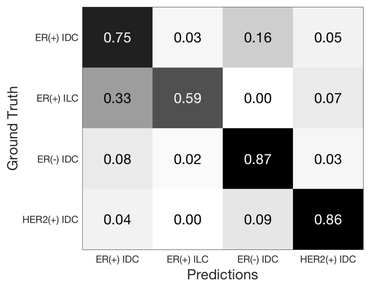
- Rotating each patch with respect to its center enlarges the dataset by a factor of $\frac{360}{\theta}$
- A typical training set contains approximately 200,000 patches
- Newly generated patches inherit the class label from the original patch
- Extra training examples enrich observations of cellular distributions

5.2 Finetuning *CaffeNet*

- *CaffeNet* is a variant of the popular *AlexNet* trained on the ImageNet dataset comprised of millions of images
- Utilize *CaffeNet* as powerful feature extractor by retraining the last fully-connected layer on IF images



6. Results

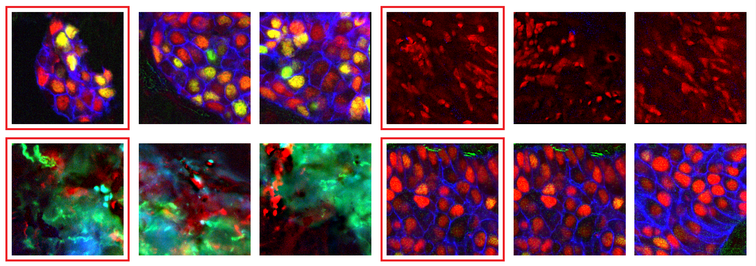


	256 × 256					512 × 512
θ step size	3	6	6	18	36	9
patch overlap	-	-	64	-	-	-
CNN Acc	0.79	0.77	**0.81**	0.76	0.75	0.78

Classification accuracy by patch size, rotation step size and overlap. Smaller θ and greater overlap increase classification accuracy.

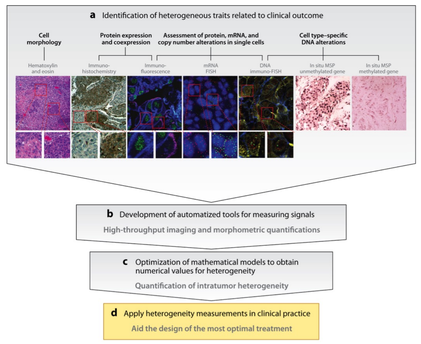
A confusion matrix of the best performing configuration is shown to the left.

7. Nearest Neighbor Visualization



Nearest neighbors using fc6 features for ER(+) IDC, ER (+) ILC, ER(-) IDC and HER2(+) IDC (left to right). In each group, the tissue patch bounded by the red box is the query image.

8. Future Directions



Spark

Lightning

- Apply method to large-scale datasets
- Accommodate data obtained from other imaging modalities such as histopathology and immunohistochemistry
- Incorporate additional scalable machine learning algorithms into the pipeline
- Expand visualization tools for use by the clinician

9. Acknowledgements & References

TECBio REU @ Pitt is supported by the National Science Foundation under Grant DBI-1263020 and is co-funded by the Department of Defense in partnership with the NSF REU program. Many thanks to Dr. Chennubhotla and his lab group.

1. Marusyk, A., et al. “Intra-tumor heterogeneity: a looking glass for cancer?.” *Nature Reviews Cancer* 12.5 (2012): 323-334. (Section 2)

2. Almendro, V., et al. ”Cellular heterogeneity and molecular evolution in cancer.” *Annual Review of Pathology: Mechanisms of Disease* 8 (2013): 277-302. (Sections 2 and 8)

3. Gerdes, Michael J., et al. “Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue.” *PNAS* 119.29 (2013): 11982-11987. (Section 4)

4. Jia, Y., et al. “Caffe: Convolutional architecture for fast feature embedding.” *arXiv preprint*:1408.5093 (2014). (Section 5)