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

Alex Andonian

¹TECBio REU @ Pitt, Department of Computational and Systems Biology University of Pittsburgh,
²Bates College, Lewiston, Maine
Motivation

Cancer mortality compared to other diseases

- 30.4m: Number of worldwide deaths per year caused by non-communicable diseases
- 7.6m: Number of worldwide deaths per year caused by cancer
- 2.5m: Good news - 1 in 3 avoidable deaths with prevention, early detection, and treatment

% fatality rate (chance of a person dying if they have the disease)

Total non-communicable disease (including cancer, cardiovascular disease, respiratory diseases, diabetes)
Tumor Heterogeneity
Data

Tissue microarray of 99 samples consisting of triplicate, 1mm diameter cores from 24 invasive breast tumor tissues.

Immunohistochemical staining revealed 4 cohorts:

- **ER(+) IDC**
- **ER(+) ILC**
- **ER(-) IDC**
- **HER2(+) IDC**

**IDC** - invasive ductal carcinoma

**ILC** - invasive lobular carcinoma
Multiplexed Immunofluorescence Imaging

A

1. Acquire Background Autofluorescence Image: 20 sec/field of view
2. Stain Slide Antigens A + B: 1 hr
3. Acquire Immunofluorescence Image: 20 sec/field of view
4. Inactivate Dye: 15 minutes
5. Acquire New Background Autofluorescence Image
6. Stain Slide Antigens C + D: 1 hr
7. Acquire New Immunofluorescence Image: 20 sec/field of view
8. Repeat (1 ... n) Times

B

A. Image Analysis
1. Register Image Stacks
2. Remove Autofluorescence
3. Image Segmentation
4. Quantitative Single Cell Feature Extraction

B. Laboratory
1. Cy5-α-Antigen A
2. Cy3-α-Antigen B
3. Cy5-α-Antigen C
4. Cy3-α-Antigen D
Research Goals

Develop high throughput informatics tools for integrating and analyzing cancer data obtained from a variety of imaging modalities

1. Cancer Classification

2. IF Signatures
Neural Networks - Biological Motivation
Convolutional Neural Networks

CNNs are very similar to ordinary neural networks, but...

- Now, we make the explicit assumption that input are images

- Since fully connected layers don’t scale well - take advantage of the fact that portions of images are correlated

Example filters learned by AlexNet
Patch Selection and Augmentation

Discriminative patch selection

Each patch is rotated 360° with step size θ
Results

Classification accuracy by configuration:

<table>
<thead>
<tr>
<th>θ step size</th>
<th>256 x 256</th>
<th>512 x 512</th>
</tr>
</thead>
<tbody>
<tr>
<td>patch overlap</td>
<td>3 6 6 18 36 9</td>
<td>- - 64 - -</td>
</tr>
<tr>
<td>CNN Acc</td>
<td>0.79 0.77 0.81 0.76 0.75 0.78</td>
<td></td>
</tr>
</tbody>
</table>

Classification accuracy:
- **increases** with patch overlap
- **decreases** with rotation step size

Confusion matrix for best performing configuration
Research Goals

Develop high throughput informatics tools for integrating and analyzing cancer data obtained from a variety of imaging modalities

1. Cancer Classification

2. IF Signatures
Nearest Neighbor Visualization

Could form the basis for a powerful and interactive visualization tool for clinicians
Conclusions

The main contributions of this work:

- classify cancer subtypes with respectable accuracy
- Identify immunofluorescent signatures associated with a cancer subtype
Future Work

a. Identification of heterogeneous traits related to clinical outcome
   - Cell morphology
     - Hematoxylin and eosin
   - Protein expression and coexpression
     - Immunohistochemistry
   - Assessment of protein, mRNA, and copy number alterations in single cells
     - Immunofluorescence
     - mRNA FISH
   - Cell type–specific DNA alterations
     - DNA immuno-FISH
     - In situ MSP unmethylated gene
     - In situ MSP methylated gene

b. Development of automatized tools for measuring signals
   - High-throughput imaging and morphometric quantifications

b. Optimization of mathematical models to obtain numerical values for heterogeneity
   - Quantification of intratumor heterogeneity

d. Apply heterogeneity measurements in clinical practice
   - Aid the design of the most optimal treatment
Acknowledgements

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.
References

Questions?