Morphologic Analysis of Glioblastoma Identifies Morphology-Driven Clusters and Molecular Correlates Associated With Patient Survival

Lee Cooper¹, Jun Kong¹, David Gutman¹,², Daniel Brat³, Joel Saltz¹,²,³

¹Center for Comprehensive Informatics
²Department of Biomedical Informatics
³Department of Pathology and Laboratory Medicine, Emory University, Emory University Hospital, Atlanta, GA
In Silico research using public data sources

Integrated Analysis
TCGA and whole slide imaging

- Scans of frozen tissue associated with molecular studies
- Scans of diagnostic-block permanent sections
- 20X magnification (40X possible?)
- Pathology evaluations (%necrosis, %tumor nuclei, histology 0,+1,+2)
Glioblastoma morphology

- Cell morphologies
- Are there clusters of GBM morphology?
- Are there morphological links to patient outcome and molecular characteristics?
Computational Pathology and Correlative Analysis

Morphology Engine
- Segmentation
- Feature Extraction
- PAIS Database
- Patient Modeling

Clustering Engine
- Normalization
- Feature Selection
- Consensus Clustering
- Multidimensional Scaling

Correlative Engine
- Survival Analysis
- Molecular Classes
- Human Pathology
- Genetic Alterations

Genome Wide Analysis
- Differential Expression
- DNA Methylation
- Copy Number Analysis
- Integrate Expression, Methylation, Genetics
- * Gene Ontology and Pathway Analyses

Examples:
- Proneural
  - Classical
  - Mesenchymal
  - Proliferative
  - GCIMP+
- TP53 +/-
  - EGFR Amp.
  - CDKN2A Del.
TABLE I
NUCLEAR FEATURES

<table>
<thead>
<tr>
<th>Category</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morphometry</td>
<td>Area, Perimeter, Eccentricity, Circularity, Major Axis Length, Minor Axis Length, Extent Ratio</td>
</tr>
<tr>
<td>Intensity Statistics</td>
<td>Mean Intensity, Max Intensity, Min Intensity, Std. Dev. Intensity</td>
</tr>
<tr>
<td>Texture</td>
<td>Entropy, Energy, Skewness, Kurtosis</td>
</tr>
</tbody>
</table>

*Note: Set of 23 features for characterization of nuclei fall into four broad categories.*
Clustering engine

Patient Morphology Profiles

Clustering Engine

Normalization
Feature Selection
Consensus Clustering
Multidimensional Scaling

Entropy
Feature Index
Correlative engine and genome wide analysis

Patient Cluster Labels

Correlative Engine

Survival Analysis

Molecular Classes
- Proneural
- Classical
- Mesenchymal
- Proliferative
- GCIMP+

Human Pathology

Genetic Alterations
- TP53 +/-
- EGFR Amp.
- CDKN2A Del.

Genome Wide Analysis

Differential Expression
DNA Methylation
Copy Number Analysis
Integrate Expression, Methylation, Genetics
* Gene Ontology and Pathway Analyses
Clustering identifies three morphological groups

- Analyzed 200 million nuclei from 162 TCGA GBMs
- Named for functions of associated genes: Cell Cycle (CC), Chromatin Modification (CM), Protein Biosynthesis (PB)
- Prognostically-significant (logrank \( p=4.5\times10^{-4} \))
Representative nuclei

CC

CM

PB
Validation

- Separate set of 84 GBMs from Henry Ford Hospital
## Associations

<table>
<thead>
<tr>
<th></th>
<th>CC Cluster</th>
<th>CM Cluster</th>
<th>PB Cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prognosis</strong></td>
<td>Average</td>
<td>Poor</td>
<td>Better</td>
</tr>
<tr>
<td><strong>Subtype Associations</strong></td>
<td>Neural Depleted</td>
<td>Neural enriched</td>
<td>Proneural Depleted</td>
</tr>
<tr>
<td><strong>Pathology</strong></td>
<td>Small cells enriched</td>
<td>Lymphocytes enriched</td>
<td>Inflammation depleted</td>
</tr>
<tr>
<td><strong>Genetics</strong></td>
<td>(NF1) mutant depleted (TP53) mutant depleted</td>
<td>None</td>
<td>None</td>
</tr>
</tbody>
</table>
Transcriptional class associations

Cluster Subtype Percentage (%)

- Classical
- Mesenchymal
- Neural
- Proneural
## Molecular associations

<table>
<thead>
<tr>
<th></th>
<th>CC Cluster</th>
<th>CM Cluster</th>
<th>PB Cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prognosis</strong></td>
<td>Average</td>
<td>Poor</td>
<td>Better</td>
</tr>
<tr>
<td><strong>Differential Expression</strong></td>
<td>2740 / 663 Genes up/down</td>
<td>200 / 463 Genes up/down</td>
<td>0 / 188 Genes up/down</td>
</tr>
<tr>
<td></td>
<td>97 / 100 miRNAs up/down</td>
<td>121 / 81 miRNAs up/down</td>
<td>15 / 5 miRNAs up/down</td>
</tr>
<tr>
<td><strong>Differential Methylation</strong></td>
<td>69 Genes hypermethylated</td>
<td>244 Genes hypermethylated</td>
<td>45 Genes hypomethylated</td>
</tr>
<tr>
<td><strong>Copy Number</strong></td>
<td>1068 Deletions</td>
<td>301 Deletions</td>
<td>399 Deletions</td>
</tr>
<tr>
<td></td>
<td>38 Amplifications</td>
<td>5 Amplifications</td>
<td>7 Amplifications</td>
</tr>
<tr>
<td><strong>Expression Mapping</strong></td>
<td>23 mapped to methylated sites</td>
<td>8 mapped to methylated sites</td>
<td>1 mapped to methylated sites</td>
</tr>
<tr>
<td></td>
<td>595 mapped to CNV sites</td>
<td>27 mapped to CNV sites</td>
<td>19 mapped to CNV sites</td>
</tr>
</tbody>
</table>
Gene Ontology and Pathway Analysis

- Nuclear lumen localization most highly enriched in cluster associated genes
  \( \text{CC } p=2.8e^{-36}, \text{ CM } p=2.17e^{-19}, \text{ PB } p=1.08e^{-15} \)

- Other enriched GO terms: DNA repair, cell cycle, protein biosynthesis, chromatin modification, m-phase

- Differences in activation of cancer-related pathways including \textit{ATM} and \textit{TP53} DNA damage checkpoints, \textit{NF} \textit{κB} pathway, \textit{Wnt} signaling and \textit{PTEN/AKT} pathways
Conclusion

• Whole-slide images contain signal
• Image analysis can provide scalable, quantitative measurements of cellular morphology
• Datasets like TCGA present a unique opportunity to correlate morphology with genomics and patient outcome
• Need more complex models to account for heterogeneity
Thank You

- Emory In Silico Brain Tumor Research Center
  NCI Contract HHSN261200800001E

Joel Saltz, MD, PhD (Director)
Dan Brat, MD, PhD (Scientific PI)
David Gutman, MD, PhD
Jun Kong, PhD
Lisa Scarpace (Henry Ford Hospital)
Tom Mikkelsen, MD (Henry Ford Hospital)