

## Introduction

### Introduction

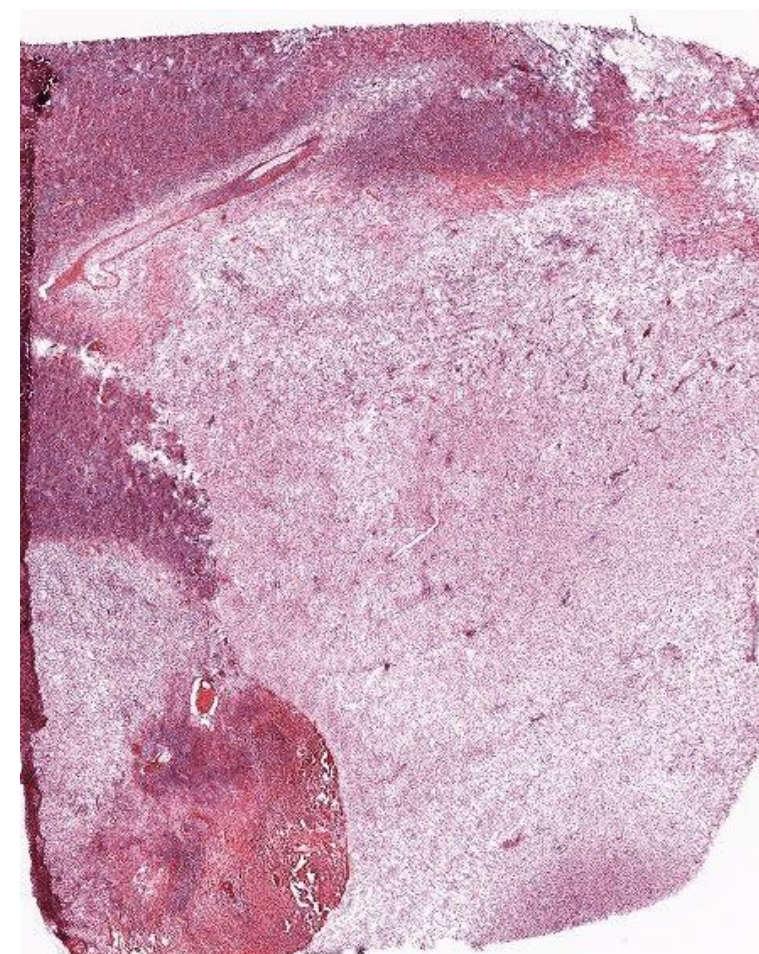
- Over the last two decades, advances in high-throughput imaging of biological tissue, and data analysis techniques, have given rise to a new field called Bioimage Informatics.<sup>[1]</sup>
- Methods in machine learning, image analysis and pattern recognition are being used for a fast, efficient and objective analysis of this data to obtain novel biological insights in areas of diagnosis, cell biology and drug discovery.
- The focus of current work is to find accurate and fast segmentation of large-scale, low resolution histological images of Glioblastoma, is a form of malignant brain tumor.
- Using various architectural features obtained from thus segmenting each image, it demonstrates the feasibility of identification of regions that are important for the diagnosis process.

## Background

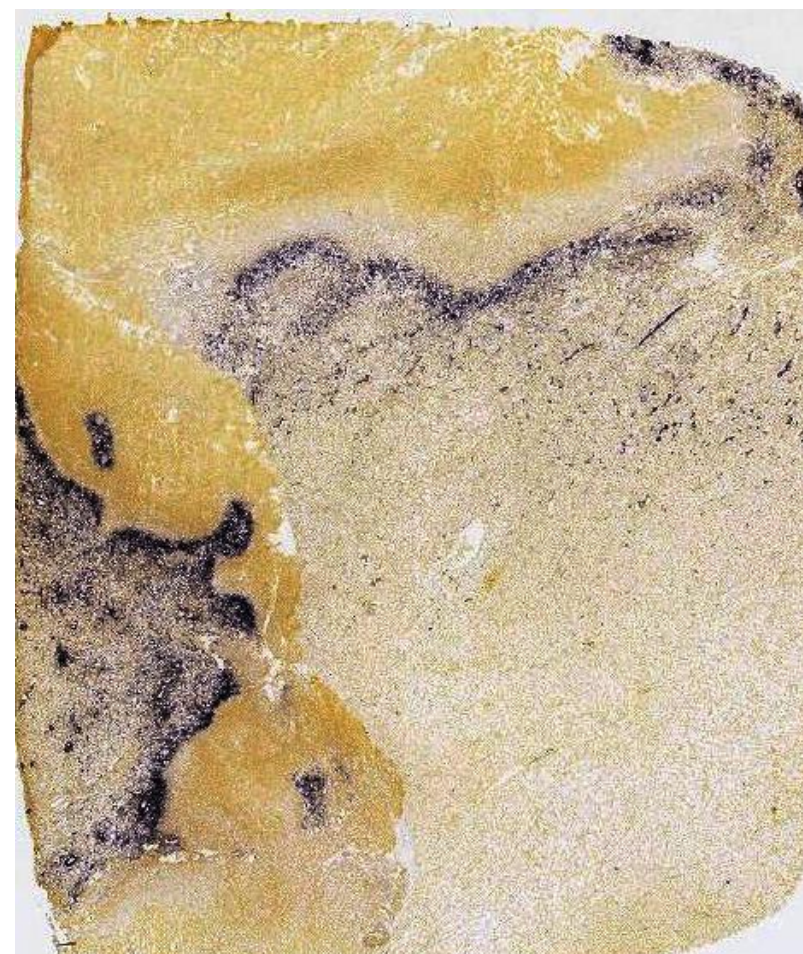
- The state-of-the-art diagnosis in clinical oncology is currently based on high resolution microscopic examination of tumours by trained pathologists.
- Due to the subjective nature of diagnosis and grading, there is significant inter-observer variability in the diagnosis of cancer.<sup>[2]</sup>
- Advances in automated imaging are generating large archives of tissue databases at a rate that is too high for traditional analysis.
- Therefore, there is a need to change the diagnostic paradigm from qualitative to quantitative.

### Ivy Glioblastoma Atlas Project<sup>[3]</sup>

- The image database made available by the Allen Brain Institute provides histological (Hematoxylin and eosin or H&E stain) and in situ hybridization (ISH) autopsy data for eight patients with glioblastoma tumors in different parts of the brain.
- The database provides over 4500 images that are 300 megapixels each, providing a wealth of information.
- This also allows us to develop methods to analyze tumor histopathology and conduct feasibility studies for computer-aided diagnosis.



H&E stain



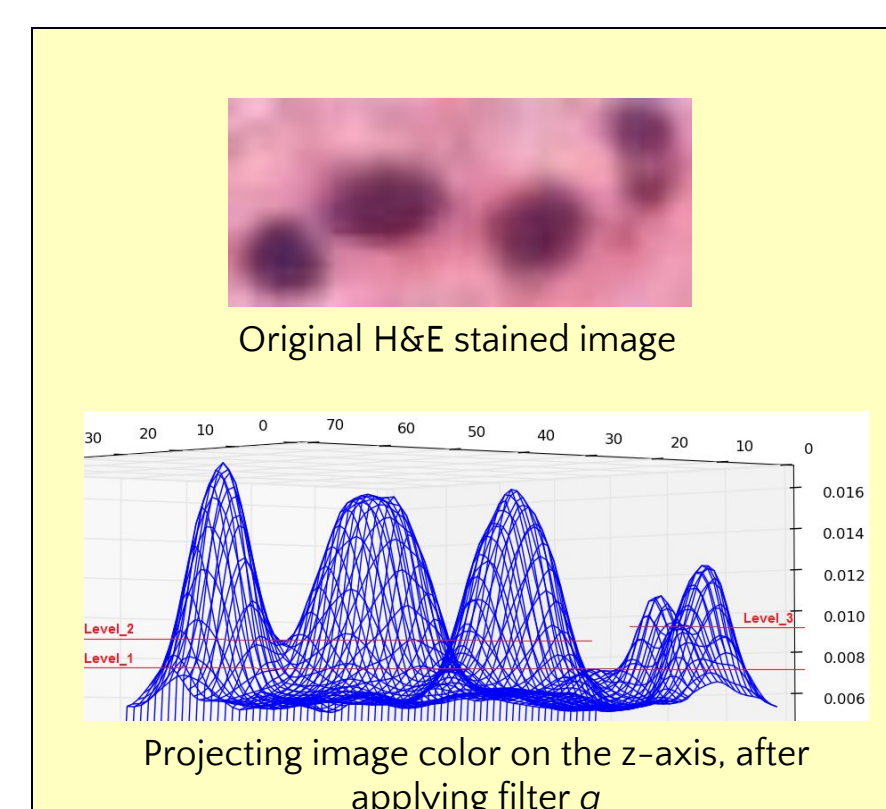
ISH stain

## Background

### Cell Segmentation

- One of the fundamental steps in CAD is the identification of histological structures within a tissue -- cell segmentation.
- The shape, size, count, location, intensity, and texture of these structures provide vital information about the presence and extent of a disease.

#### Iterative Thresholding



**Iterative thresholding**, a color-based thresholding method, finds a threshold at multiple levels until all the objects in the image are found.

- Apply the following filter,  $g$ , to the image:

$$g = 1/(1 + G * I),$$

where  $G$  is a gaussian kernel and  $I$  the image

A **subregion**  $s$  is defined as a continuous section of non-zero pixels in a given region  $r$  of an image.

#### Algorithm:

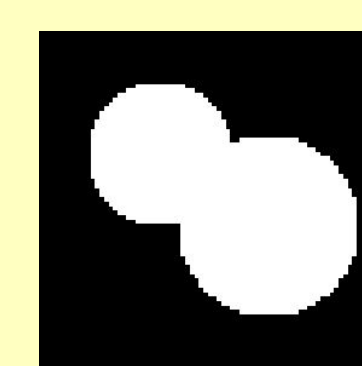
- Initial call:**  $Segment(I)$
- Function  $Segment(r)$ :**

$$r_{new} = r - \{mean(r) + std_{dev}(r)\}$$

If  $countSubregion(r_{new}) > 1$ :  
For each  $s$  in  $getSubregion(r_{new})$ :  
 $Segment(s)$

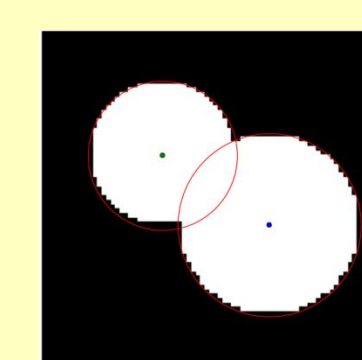
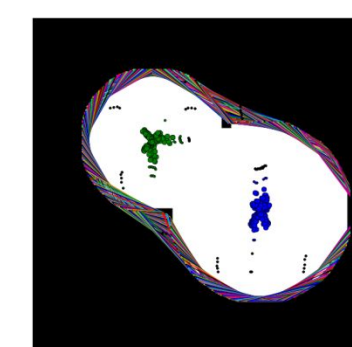
- This method is sensitive to local variations in shades and hues, unlike standard global thresholding methods.
- Works especially well in low resolution images, where sophisticated methods, such as level set and countour detection fail.
- Fast and efficient when there are thousands of objects to be detected within seconds.

#### Cleaving



In case of adjacent objects/cells with no apparent boundary such as shown above, how do we segment and identify the number of objects present?

If multiple triangles are inscribed within this shape, most of the **circumcircles** for these triangles will coincide with either of the constituent circles and form **clusters** of circumcenters.



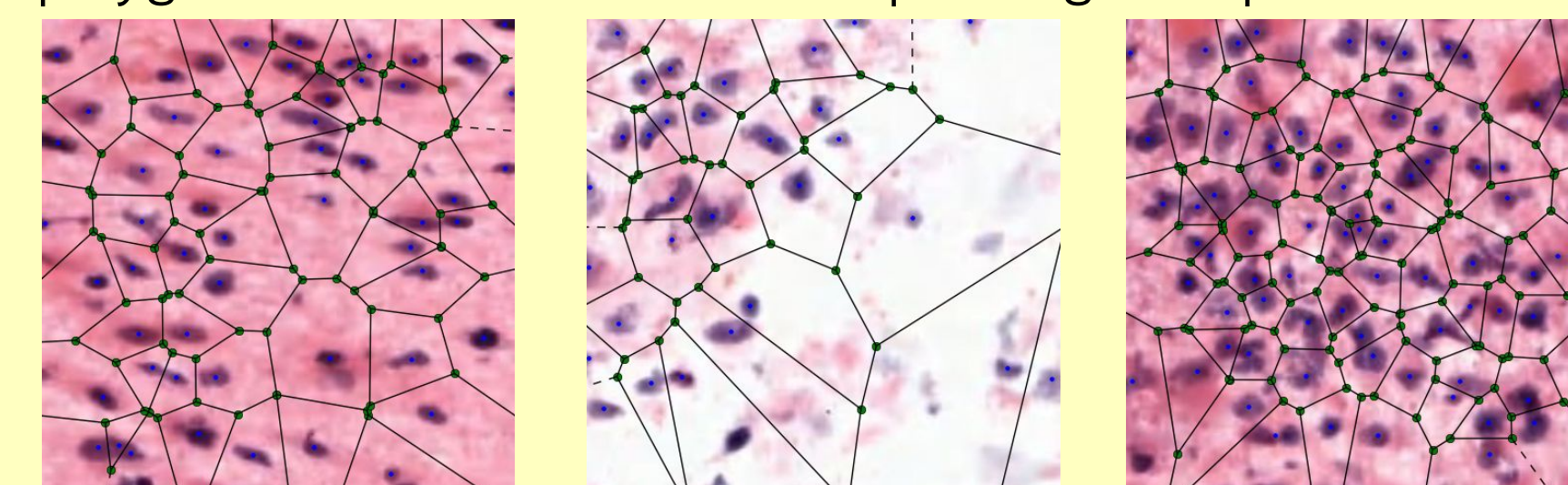
The **cluster** center and the average **circumradius** of triangles in each cluster give the approximate center and boundary of the object. The method so developed is called **cleaving**.

### Feature Extraction

- Topological features – which describe the spatial arrangement of nuclei in a histopathology image provides important information about brain tumors.<sup>[4,5]</sup>
- They are derived by constructing graphs with the nuclei as nodes, and extracting metrics such as the vertex degree, connectivity, and eccentricity as features.

#### Voronoi Tessellation

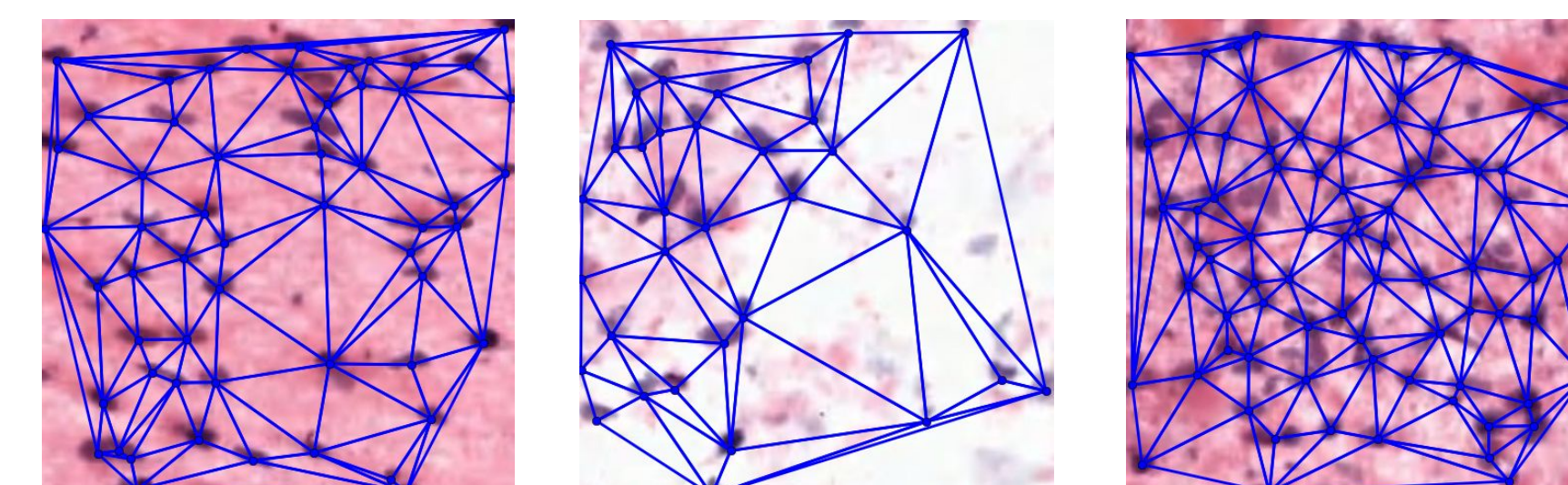
A voronoi tessellation partitions a space containing  $n$  seed points, into  $n$  convex polygons such that each point in a polygon is closest to the its corresponding seed point.



Features extracted: Area of polygon, roundness of polygon, disorder of areas, number of neighbours

#### Delaunay Triangulation

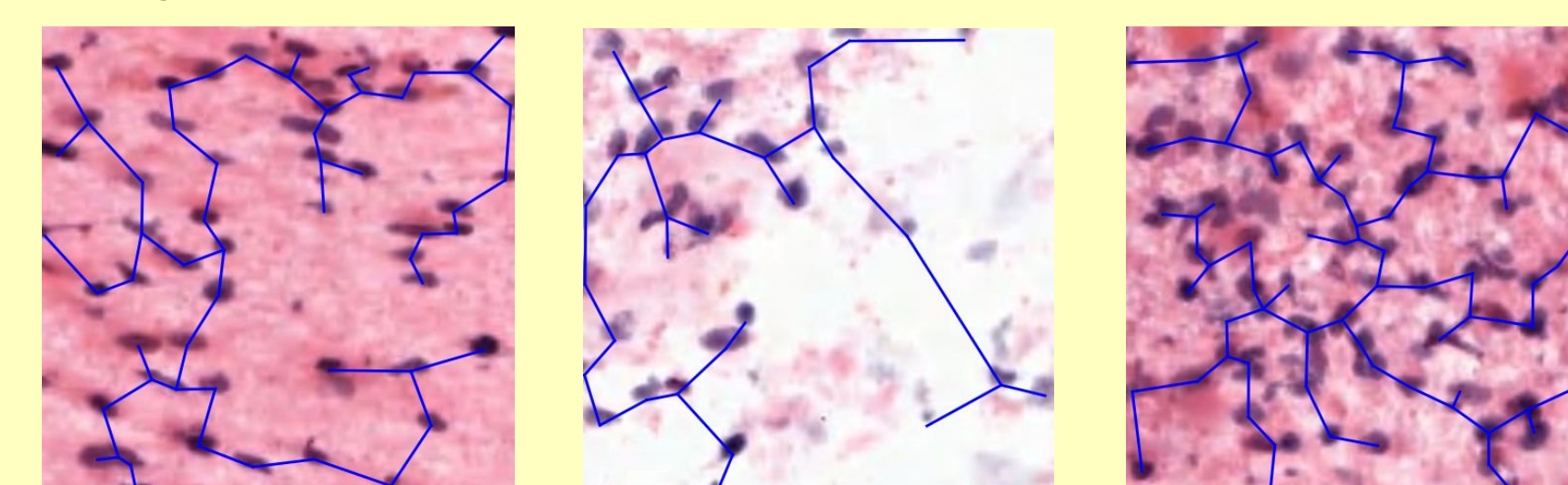
A delaunay triangulation, is a triangulation over  $n$  points on a plane such that no point  $P$  is inside the circumcircle of any triangle, and the minimum angle of all the triangles are maximized.



Features extracted: Average edge length, maximum edge length, average length of edges that collect a nucleiod to its corresponding neighbours

#### Minimum Spanning Tree

A minimum spanning tree is a tree constructed over a graph such that the path between any two nodes is the shortest of all possible paths. In this case, it is calculated over the triangulation described above.



Features extracted: Average length of edges, edge disorder

#### Miscellaneous features:

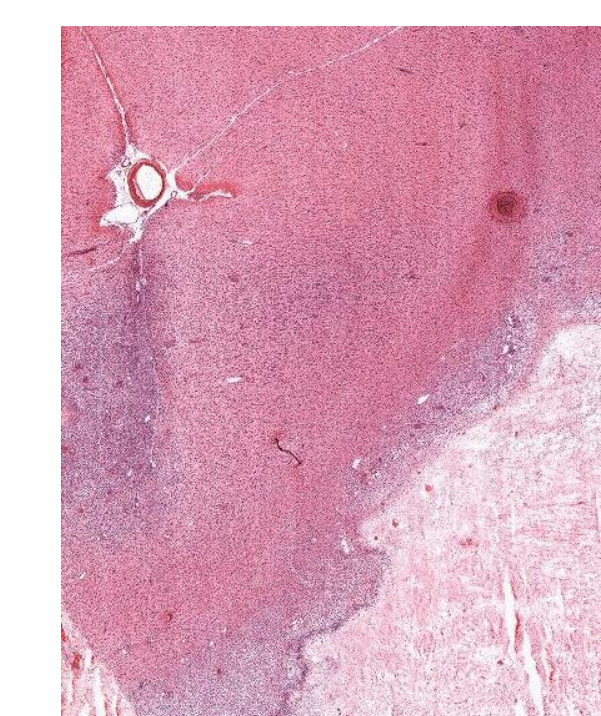
- Number of cells.
- Average grayscale intensity of the image.
- Percentage of are occupied by the cells in each image.

### Results

- The 300 megapixel image is split into 256 256 pixel tiles.
- We perform cell segmentation of each image with high accuracy:

	Proprietary	OTSU	Adaptive OTSU
True Positive	94.7%	74.8%	91.6%
False positive	8.2%	16.7%	32.1%

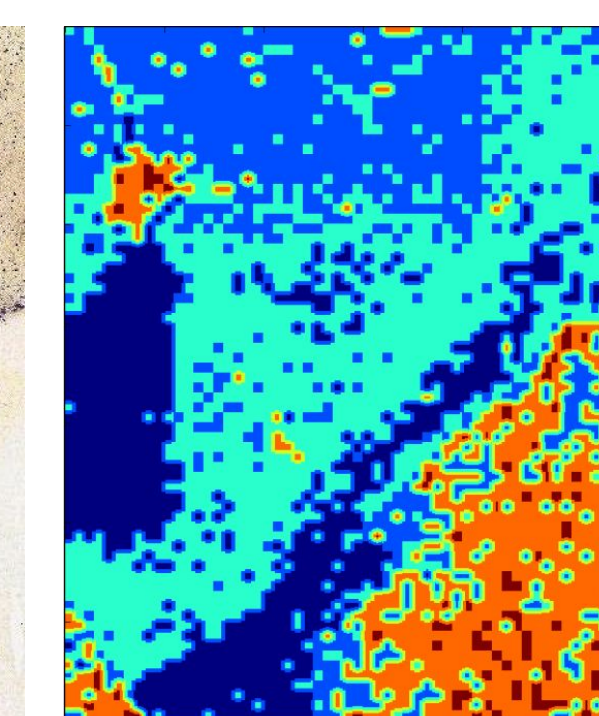
- K-means clustering is used to cluster each image based on the topological features extracted to obtain regions of interest in the image, such as tumors, white matter, necrosis and healthy tissue.
- We find a strong correlation between the clusters identified and the expression of genes corresponding to growth of tumors, indicating that robust classifiers can be developed to identify Glioblastoma tumors.



Original H&E Image



ISH stain for TIMP1, indicating Glioblastoma tumor



Identified regions of interest, clustering with K=8

## Conclusion and Future Work

- The cell segmentation and feature extraction pipeline developed here provide a fast and an efficient method to quantitatively featurize and analyze image data, thus removing inter-observer variability.
- In the immediate future, this work could be extended to develop a screening framework that identifies regions of interest for expert review.
- By introducing supervised machine learning techniques such as support vector machines and neural networks to the framework, a robust system can be developed to identify different grades of brain tumor.

## Bibliography

- Hanchuan Peng, Alex Bateman, Alfonso Valencia, and Jonathan D Wren. Bioimage informatics: a new category in bioinformatics. Bioinformatics, 28(8):10571057, 2012.
- Janet M Bruner, Lila Inouye, Gregory N Fuller, and Lauren A Langford. Diagnostic discrepancies and their clinical impact in a neuropathology referral practice. Cancer, 79(4):796803, 1997.
- Allen Brain institute. Ivy glioblastoma atlas project, 2012.
- Cigdem Demir, S Humayun Gultekin, and Bulent Yener. Learning the topological properties of brain tumors. IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB), 2(3):262270, 2005.
- Shivang Naik, Scott Doyle, Michael Feldman, John Tomaszewski, and Anant Madabhushi. Gland segmentation and computerized gleason grading of prostate histology by integrating low-, high-level and domain specic information. In MIAAB Workshop. Citeseer, 2007.