



caBIG™ cancer Biomedical  
Informatics Grid™

an initiative of the National Cancer Institute

# Integrative Multi-Scale Biomedical Informatics

*Joel Saltz MD, PhD*

*Director Center for Comprehensive  
Informatics*

*Chair Department of Biomedical  
Informatics*



EMORY  
UNIVERSITY

# INTEGRATIVE BIOMEDICAL INFORMATICS ANALYSIS

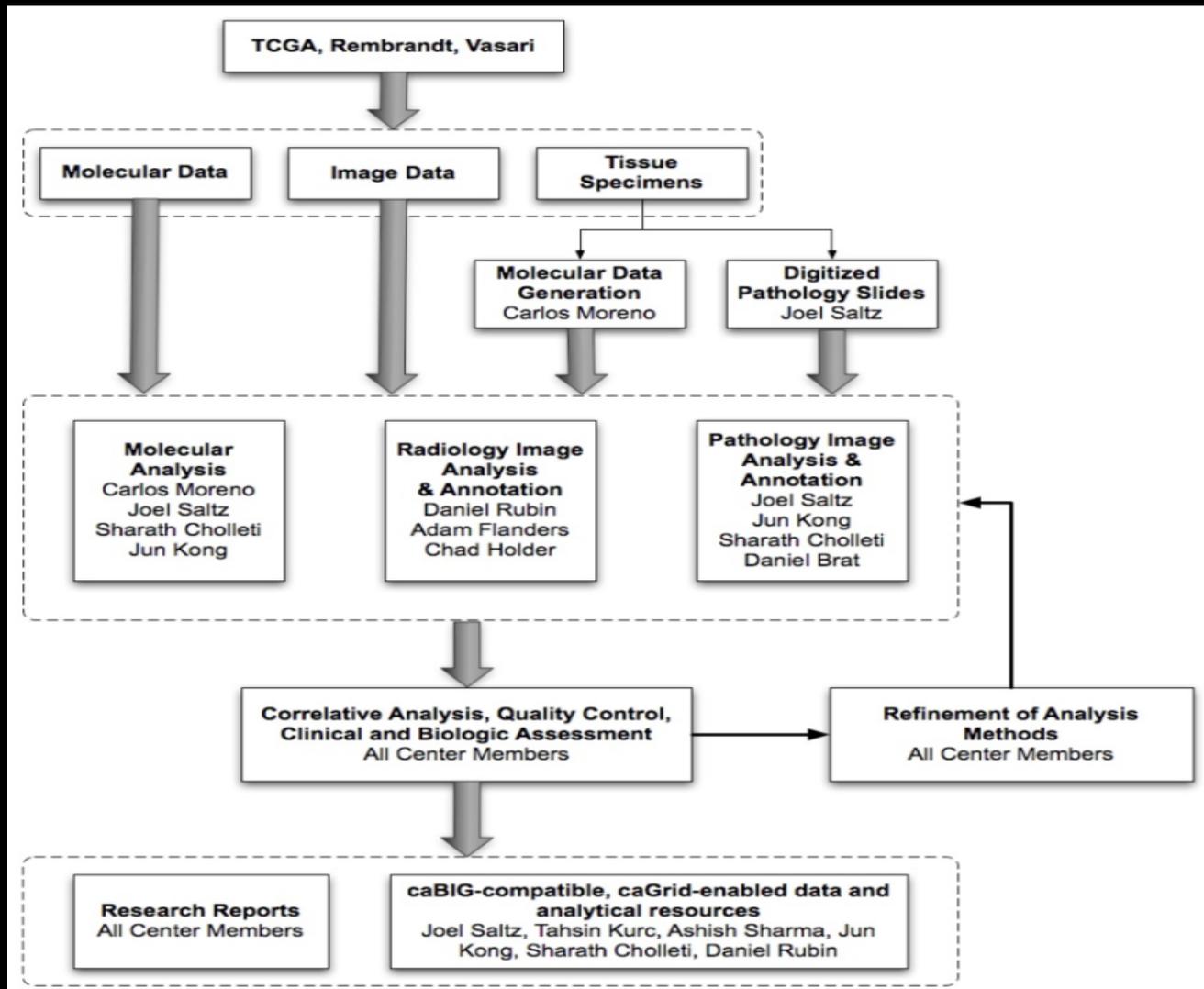
Reproducible anatomic/functional characterization at gross level (Radiology) and fine level (Pathology)

Integration of anatomic/functional characterization with multiple types of “omic” information

Create categories of jointly classified data to describe pathophysiology, predict prognosis, response to treatment



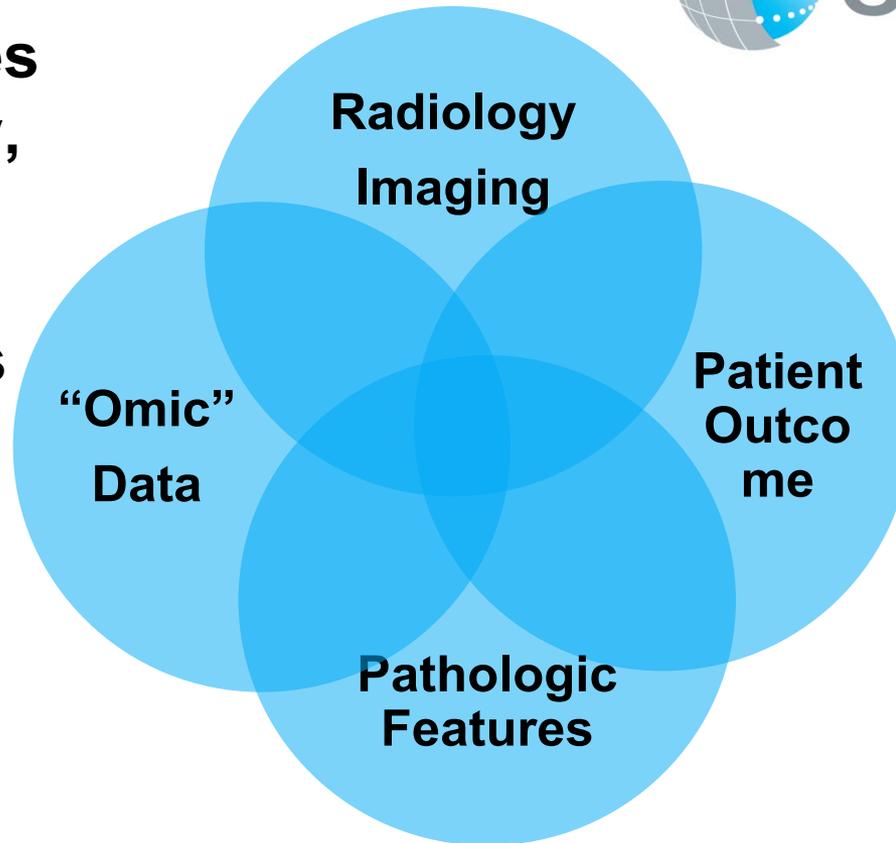
# caBIG In Silico Center for Brain Tumor Research



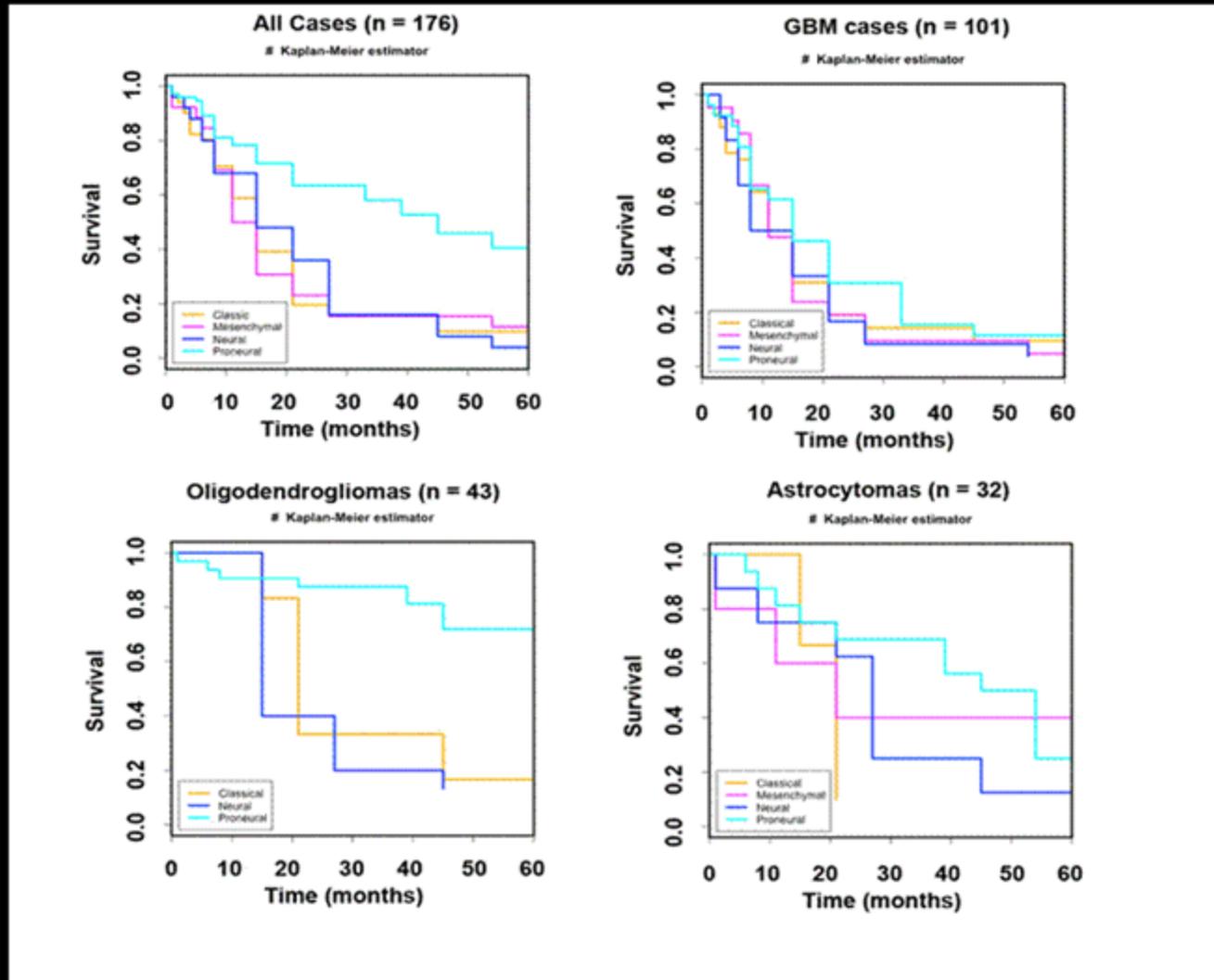
# Integration of heterogeneous multiscale information



- **Coordinated initiatives Pathology, Radiology, “omics”**
- **Exploit synergies between all initiatives to improve ability to forecast survival & response.**



# Example: Pathology and Gene Expression Joint Predictors of Recurrence/Survival



Lee Cooper Carlos Moreno

# ***In Silico* Center for Brain Tumor Research**

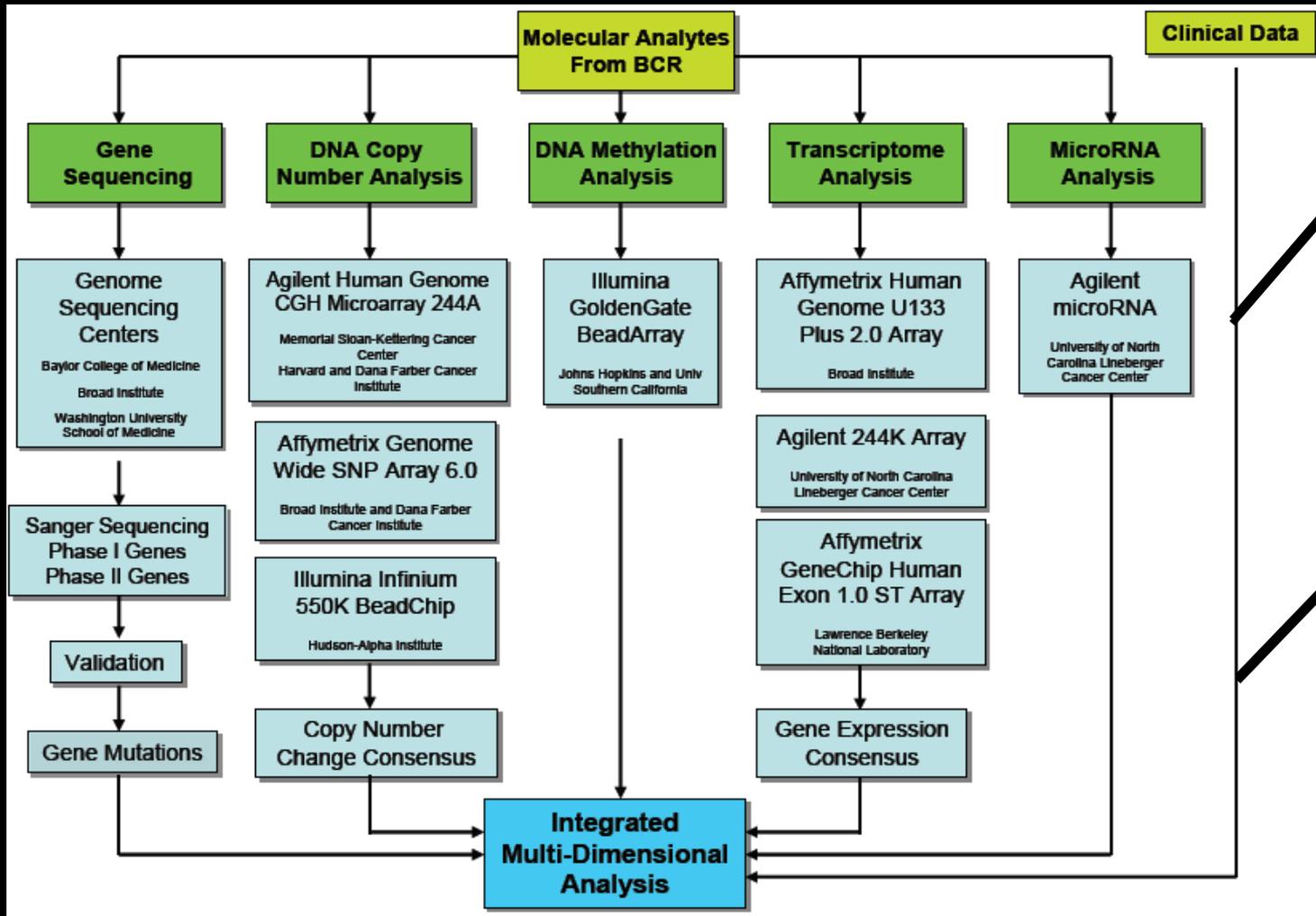
## **Key Data Sets**

REMBRANDT: Gene expression and genomics data set of all glioma subtypes

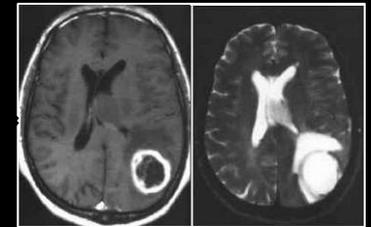
The Cancer Genome Atlas (TCGA): Rich “omics” set of GBM, digitized Pathology and Radiology

Pathology and Radiology Images from Henry Ford Hospital, Emory, Thomas Jefferson U, MD Anderson and others

# TCGA Research Network

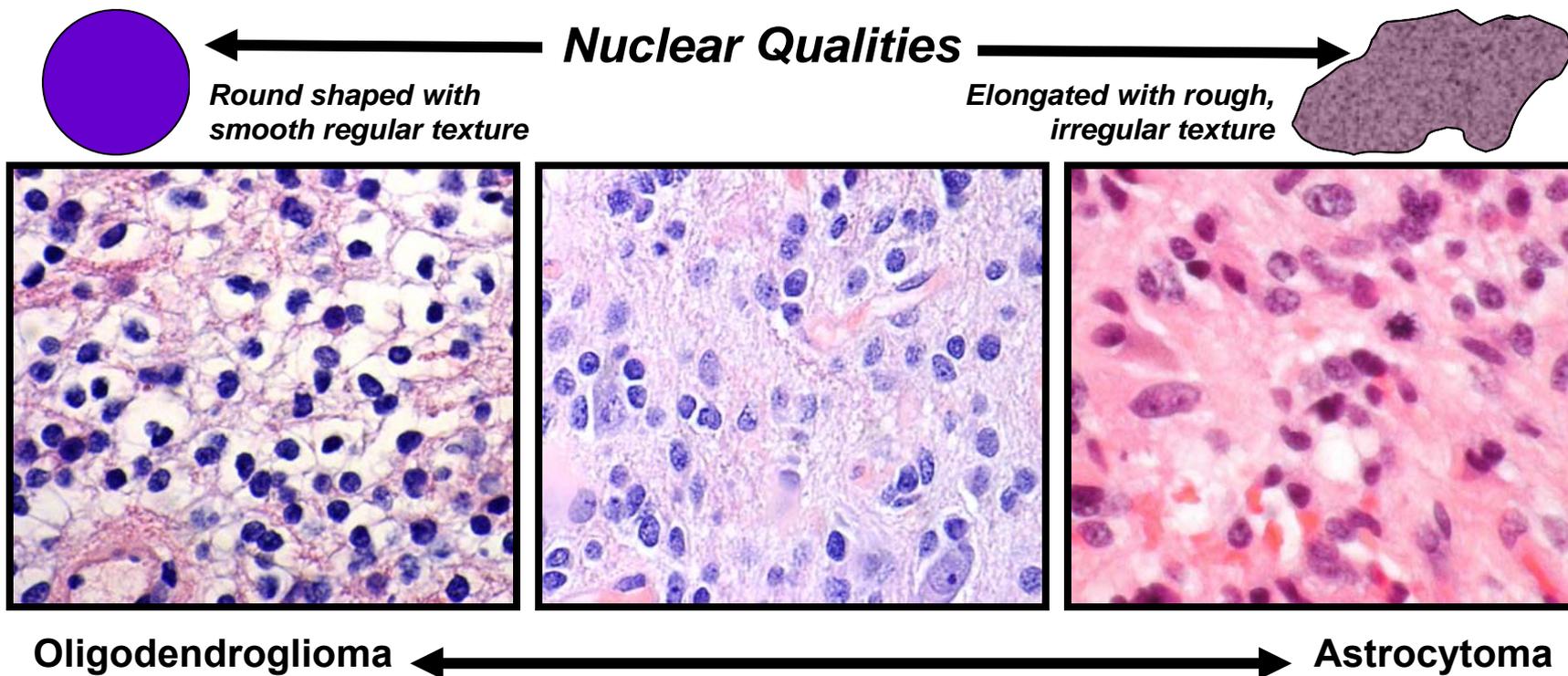


Digital Pathology



Neuroimaging

# Characterizing Pathology of Gliomas



- Pathologists and image analysis algorithms describe, segment and classify microanatomic features in whole slide images

# TCGA Brain Pathology Criteria

## Attributes that Relate to Entire Specimen

Roughly 120 TCGA patients; Three Reviewers with Dan Brat adjudicating

**Not Present:** Not detected on any block

**Present:** detected on any block

**Abundant:** present in  $\geq 50\%$  of 10x fields in  $\geq 50\%$  blocks

- Microvascular hyperplasia elements (1,2)
- Complex/glomeruloid
- Circumferential endothelial hyperplasia
- 
- Necrosis elements (3,4)
- Multiple serpentine pseudoepithelioid pattern
- Zonal necrosis
- Small cell component
- Gemistocytes
- “Oligodendroglioma-like” component with perinuclear cytoplasmic halos
- Perineuronal and/or perivascular satellitosis
- Multi-nucleated/giant cells
- Epithelial metaplasia
- Mesenchymal metaplasia
- Entrapped gray matter
- Entrapped white matter
- Micro-mineralization
- 
- Inflammation
- Macrophage/histiocytic infiltrates
- Lymphocytic infiltrates
- Polymorphonuclear leukocytic infiltrates



# TCGA Whole Slide Images

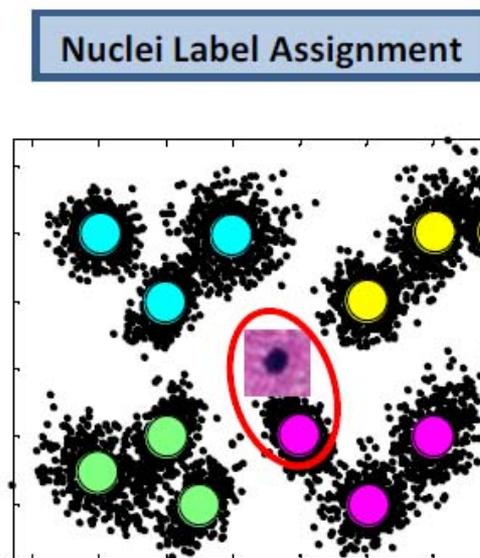
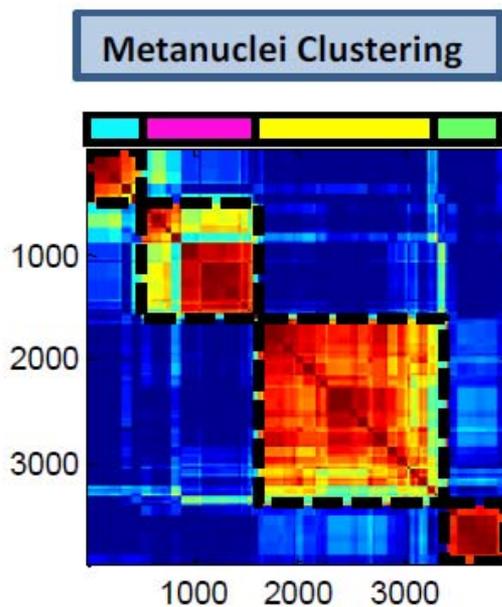
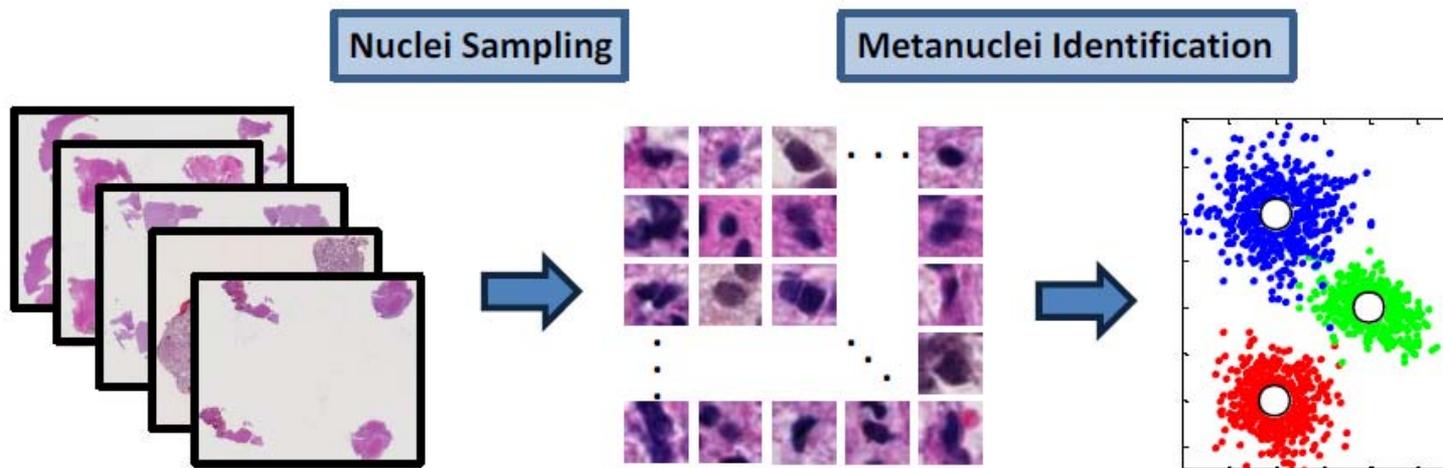
## Feature Extraction



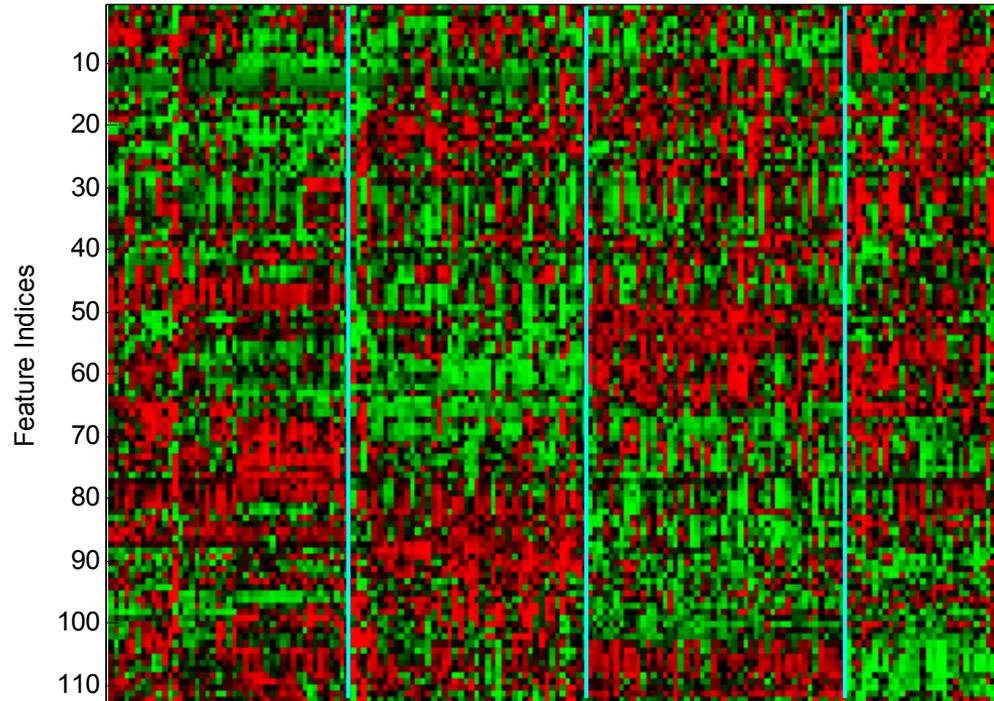
Jun Kong



# Physical Interpretations of Image Analysis Results



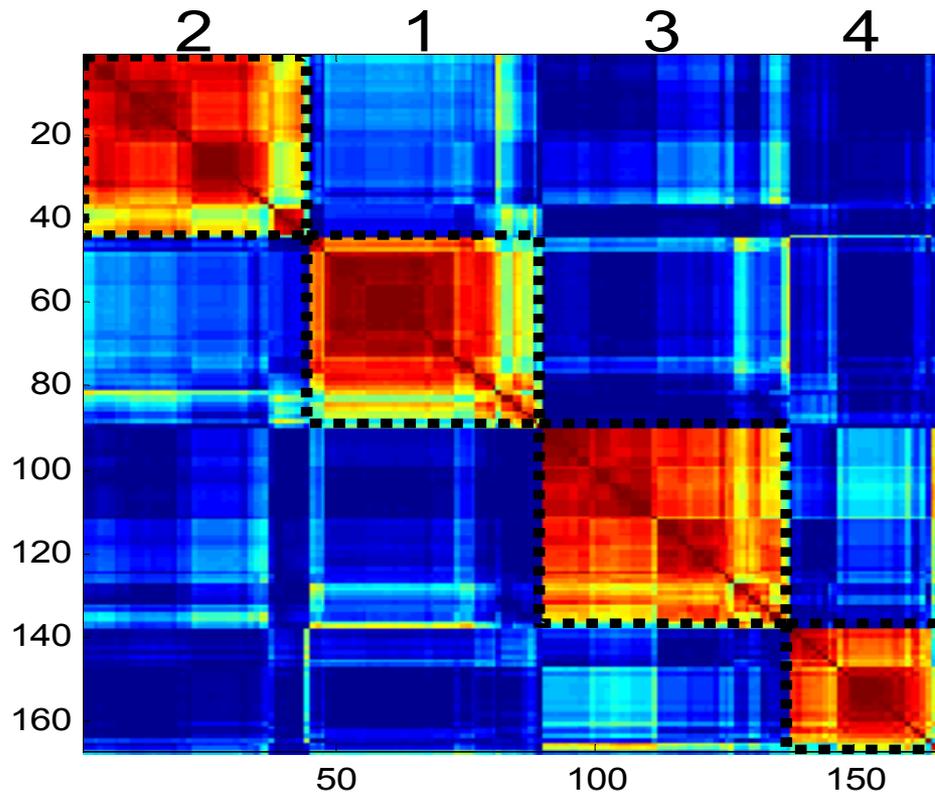
# Nuclear Features Used to Classify GBMs



Clustergram of selected features used in consensus clustering



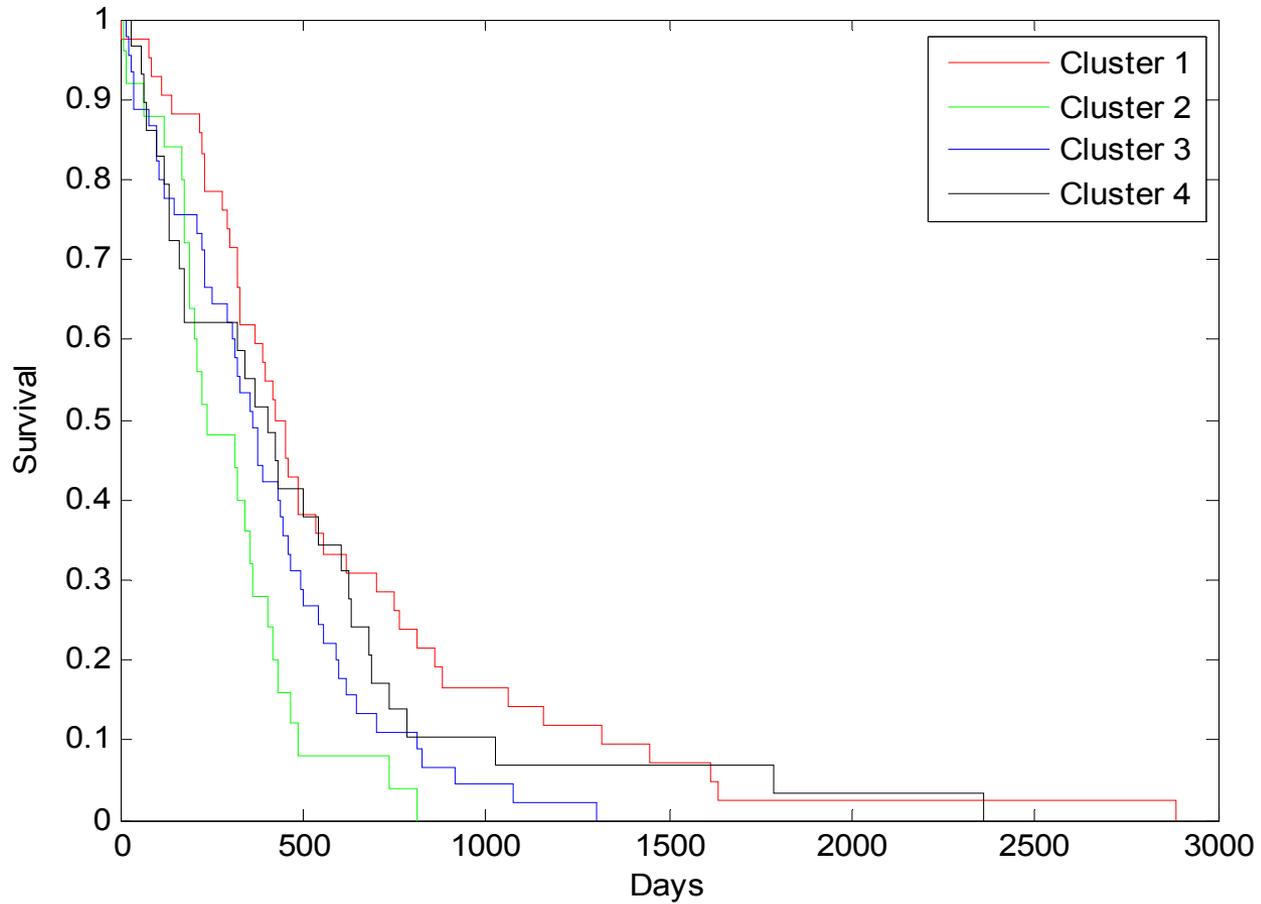
# Nuclear Features Used to Classify GBMs



## Consensus clustering of morphological signatures

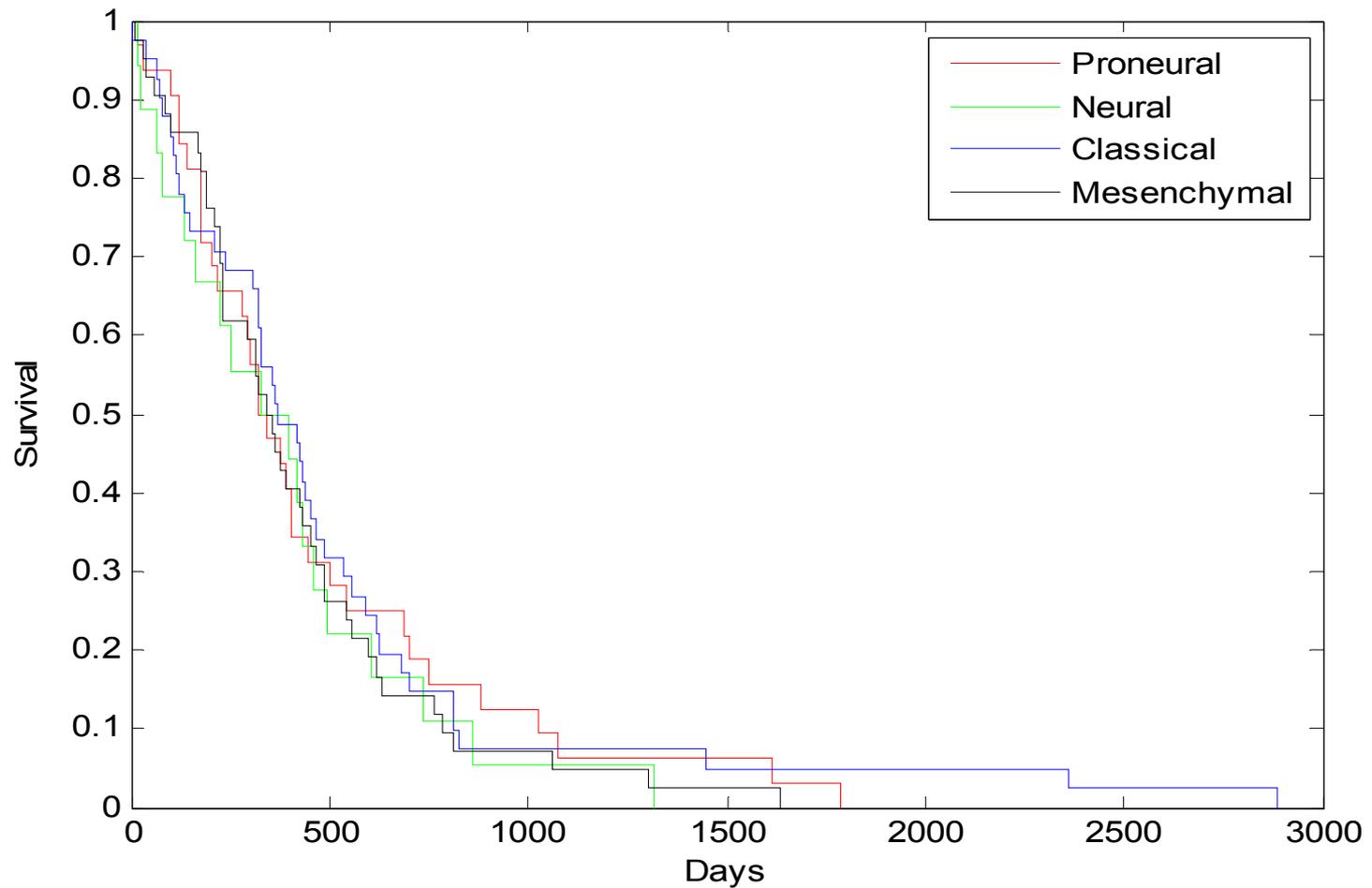
Study includes 200 million nuclei taken from 480 slides corresponding to 167 distinct patients.





## Survival of morphological clusters



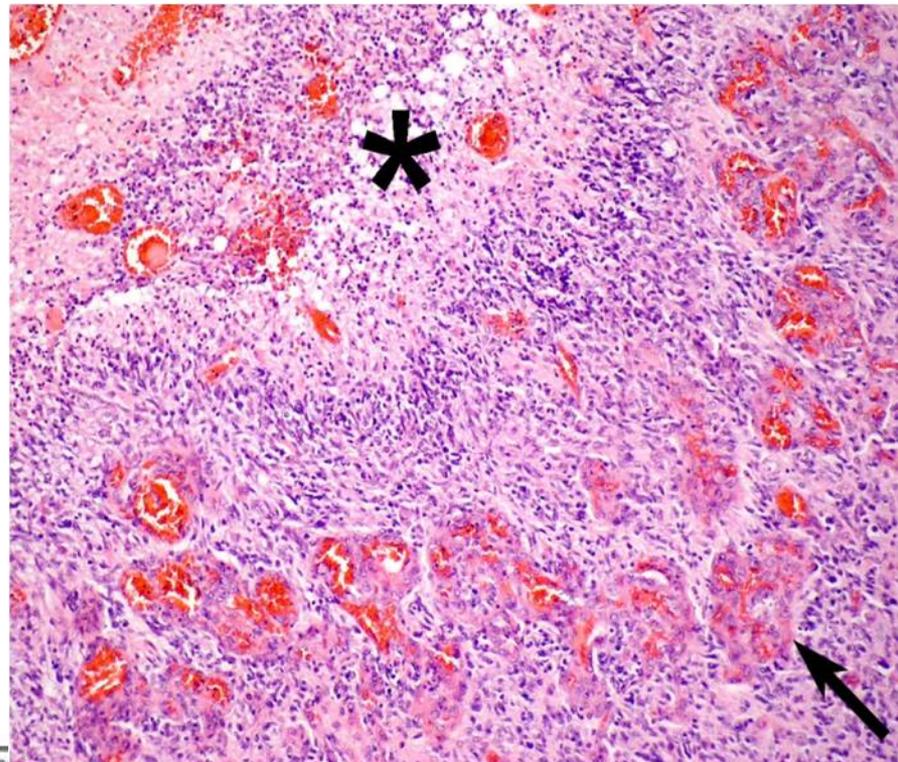


## Survival of patients by molecular tumor subtype



# Correlation of Necrosis, Angiogenesis, “omics” and Outcome

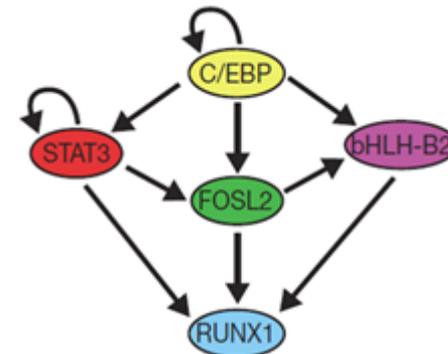
- GBMs display variable and regionally heterogeneous degrees of necrosis (asterisk) and angiogenesis
- These factors may impact gene expression profiles



## Genes Correlated with Necrosis include Transcription Factors Identified as Regulators of the Mesenchymal Transition

- Frozen sections from 88 GBM samples marked to identify regions of necrosis and angiogenesis
- Extent of both necrosis and angiogenesis calculated as a percentage of total tissue area

Gene Symbol	SAM q-value (Corrected p-value)
C/EBPB	< 0.000001
C/EBPD	< 0.000001
FOSL2	< 0.000001
STAT3	0.0047
RUNX1	0.0082



Carro MS, et al. Nature 263: 318-25, 2010



# Imaging Predictors of survival and molecular profiles in the TCGA Glioblastoma Data set

## The TCGA glioma working group

<sup>1</sup>Emory University Hospital, Atlanta, GA <sup>2</sup>National Cancer Institute, Bethesda, MD. <sup>3</sup>Thomas Jefferson University Hospital, Philadelphia, PA. <sup>4</sup>Henry Ford University Hospital, Detroit, Michigan. <sup>5</sup>National Institute of Health, Bethesda, MD. <sup>6</sup>Boston University School of Medicine, Boston, MA. <sup>7</sup>SAIC-Frederick, Inc., Frederick, MD. <sup>8</sup>University of Virginia, Charlottesville, VA. <sup>9</sup> Northwestern University Chicago, IL

### Emory

David A Gutman<sup>1</sup>

Lee Cooper<sup>1</sup>

Scott N Hwang<sup>1</sup>

Chad A Holder<sup>1</sup>

Doris Gao<sup>1</sup>

Carlos Moreno<sup>1</sup>

Arun Krishnan<sup>1</sup>

Jun Kong<sup>1</sup>

Seena Dehkharghani<sup>1</sup>

Joel Saltz<sup>1</sup>

Dan Brat<sup>1</sup>

### TJU/CBIT/NCI

Adam Flanders<sup>3</sup>

Eric Huang<sup>2</sup>

Robert J Clifford<sup>2</sup>

Dina Hammoud<sup>3</sup>

John Freymann<sup>7</sup>

Justin Kirby<sup>7</sup>

Carl Jaffe<sup>6</sup>

### UVA/Northwestern

Max Wintermark<sup>8</sup>

Manal Jilwan<sup>8</sup>

Prashant Raghavan<sup>8</sup>

Pat Mongkolwat<sup>9</sup>

### Henry Ford

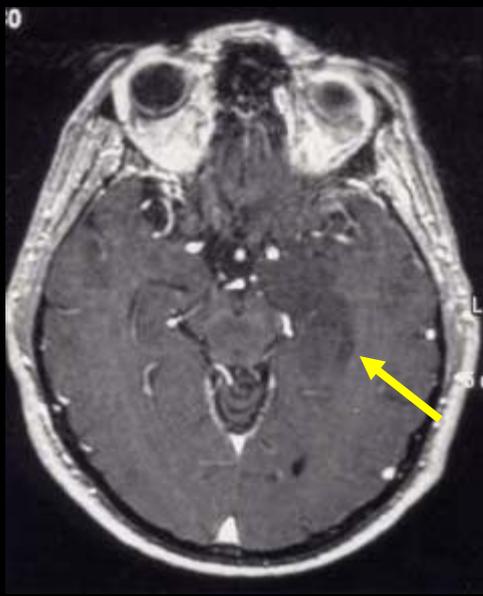
Lisa Scarpace<sup>4</sup>

Tom Mikkelsen<sup>4</sup>

# ***GBM Imaging Biomarkers***

## ***Image Quantification***

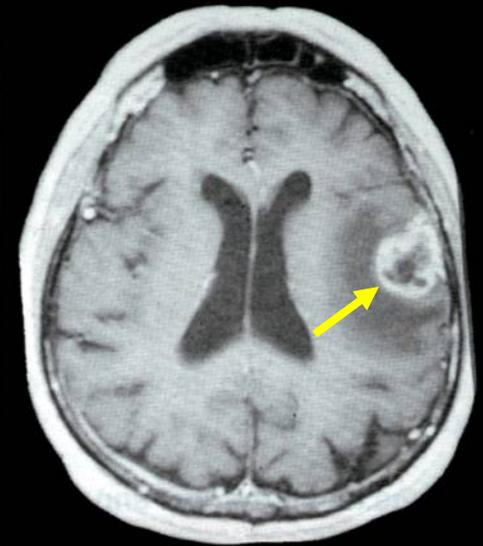
### ***Correlation with Pathology, “omics”***



No enhancement  
Normal Vessels  
Stable lesion



?



Rim-enhancement  
Vascular Changes  
Rapid progression

# Feature Sets in Neuro Radiology

Adam Flanders (TJU), Daniel Rubin (Stanford), Lori Dodd (NCI), Eric Huang (NCI), many others

- Standardized validated feature sets
- Standard terminology:
  - Comprehensive, reproducible set of imaging features of cancer
  - Quantitative, reproducible basis for assessing baseline disease and treatment response
  - Gutman (Emory) leading effort to develop complementary algorithmic pipeline
- caBIG tools NBIA and AIME data service employed to manage images, annotations
- caBIG AIM template and ClearCanvas software supports Radiologist annotation process



# AIM and AIM Template on ClearCanvas Workstation

The screenshot displays the ClearCanvas Workstation 3.0 interface. The main window shows an axial T1-weighted MRI scan of a brain with two yellow dashed lines indicating tumor measurements: 4.28 cm and 3.20 cm. The interface includes a menu bar (File, Edit, View, Tools, Annotation, Help), a toolbar with various icons, and a sidebar with 'Available AIM Annotations' and 'AIM Annotation' panels. The 'AIM Annotation' panel is active, showing a dropdown menu for 'AIM Template' set to 'Brain Tumor Measurement TCGA(VASARI-3)'. The 'Annotation Name' field contains 'TCGA-06-0164\_trr\_2011-04-01 17:54 PM'. The 'Lesion Location' section includes checkboxes for '01 - Tumor Location (F1)', '02 - Side of Tumor Epicenter (F2)', and '03 - Eloquent Brain (F3)'. The 'Morphology of Lesion Substance' section includes dropdown menus for '04 - Enhancement Quality (F4)', '05 - Proportion Enhancing (F5)', '06 - Proportion nCET (F6)', '07 - Proportion Necrosis (F7)', and '08 - Proportion of Edema (F14)'. A red arrow points to the 'Importing an AIM template' button in the toolbar. Another red arrow points to the 'Show Study' button in the 'Available AIM Annotations' panel, with the text 'An existing AIM instance' below it.

125307 - TCGA-06-0164 [AIM TCGA on ClearCanvas Workstation 3.0 TCGA]

File Edit View Tools Annotation Help

Available AIM Annotations

TCGA-06-0164\_trr\_2011-03-28 13:21 PM 4.28 cm

TCGA-06-0164\_trr\_2011-03-28 13:21 PM 3.20 cm

TCGA-06-0164 125307

AXIAL T1 SI GAD

12-Jan-1885

Acq: -

Se: 10/13

Im: 3/4

Loc: H34.0

256 x 192

ET: 0

TR: 500.00 ms

TE: 14.00 ms

HEAD

Thk: 15.0 mm

Zoom: 2.87x

W:976 L:488 (Auto)

15000.0T

BAY1OC0

DFOV: 24.0 x 24.0 cm

Importing an AIM template

AIM Annotation

AIM Template: Brain Tumor Measurement TCGA(VASARI-3)

Annotation Name: TCGA-06-0164\_trr\_2011-04-01 17:54 PM

Lesion Location

01 - Tumor Location (F1)

Frontal lobe

Temporal lobe

Insula

Parietal lobe

02 - Side of Tumor Epicenter (F2)

03 - Eloquent Brain (F3)

No Eloquent Brain

Speech Motor

Speech Receptive

Motor

Lesion

Morphology of Lesion Substance

04 - Enhancement Quality (F4)

05 - Proportion Enhancing (F5)

06 - Proportion nCET (F6)

07 - Proportion Necrosis (F7)

08 - Proportion of Edema (F14)

Create Annotation

User Info

An existing AIM instance



# Examples of Recent Findings Relating Radiology, Pathology and “Omics”

- Fisher’s exact test demonstrated an association between the presence of CDKN2A homozygous deletion and the identification of an ***ill-defined nonenhancing tumor margin*** (p=0.007).
- TP53 mutant tumors had a ***decreased mean tumor size*** (p=0.002), measured as the maximum tumor dimension in the T2-weighted or FLAIR images
- Significant association between ***minimal enhancing tumor*** (≤5% proportion of the overall tumor) and Proneural classification (p=0.0006).
- Significant association between a ***>5% proportion of necrosis*** and the ***presence of microvascular hyperplasia*** in pathology slides (p=0.008).



# Data Models to Represent Feature Sets and Experimental Metadata



## *PAIS |pās| : Pathology Analytical Imaging Standards*

- Provide semantically enabled data model to support pathology analytical imaging
- Data objects, comprehensive data types, and flexible relationships
- Object-oriented design, easily extensible
- Reuse existing standards
  - Reuse relevant classes already defined in AIM
  - Follow DICOM WG 26 metadata specifications on WSI reference
  - Specimen information in DICOM Supplement 122 and caTissue
  - Use caDSR for CDE and NCI Thesaurus for ontology concepts

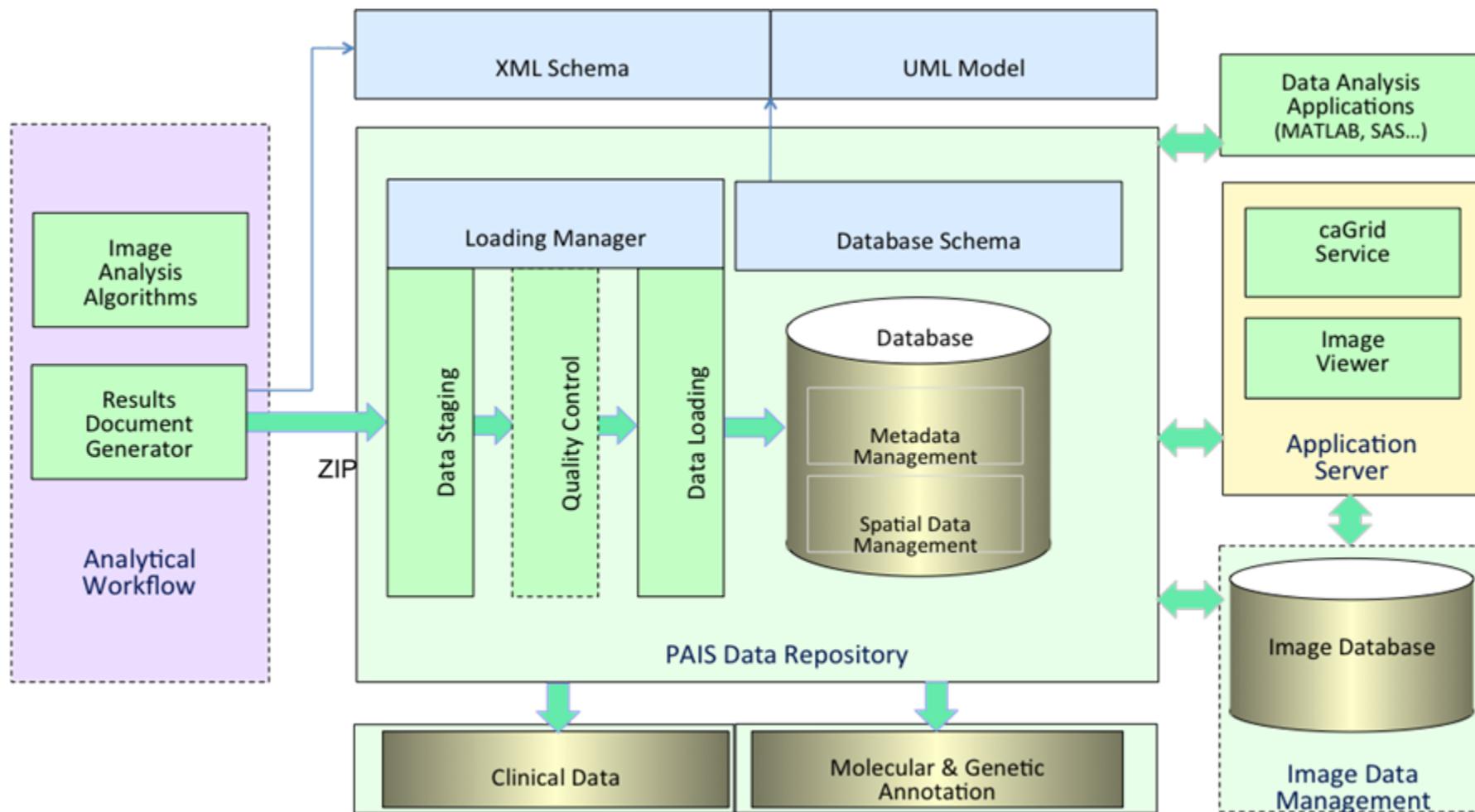
# PAIS Database



- Implemented with IBM DB2 for large scale pathology image metadata (~million markups per slide)
- Represented by a complex data model capturing multi-faceted information including markups, annotations, algorithm provenance, specimen, etc.
- Support for complex relationships and spatial query: multi-level granularities, relationships between markups and annotations, spatial and nested relationships

**Fusheng Wang, Emory**

# Feature Derivation, Analysis and Management Framework



# Pathology Imaging GIS

Fusheng Wang

Center for Comprehensive Informatics

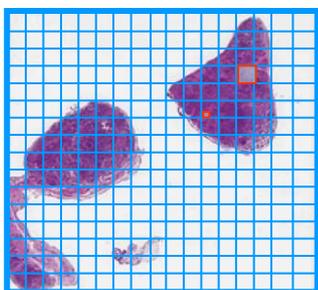
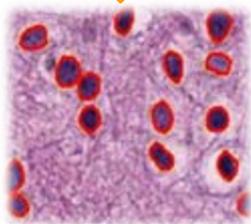


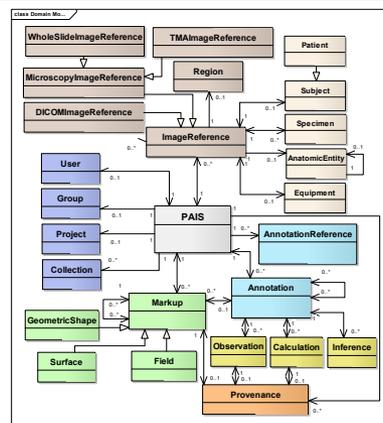
Image analysis



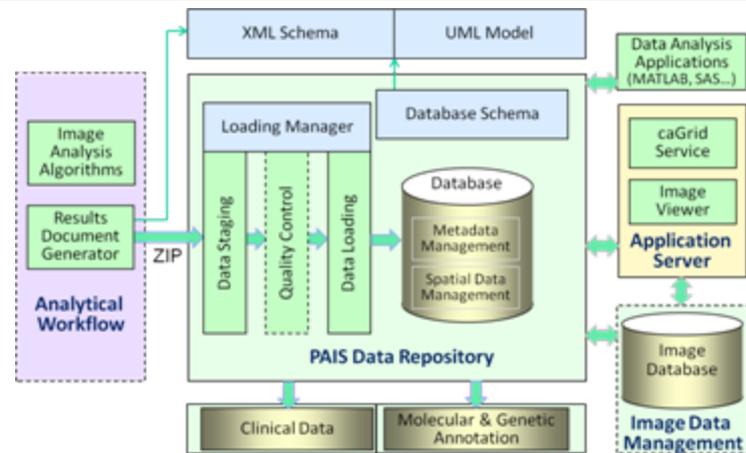
Segmentation

Feature	Value
Area	723
Perimeter	24.5
Circularity	0.26
Median	230
Variance	17.8
...	...

Feature extraction

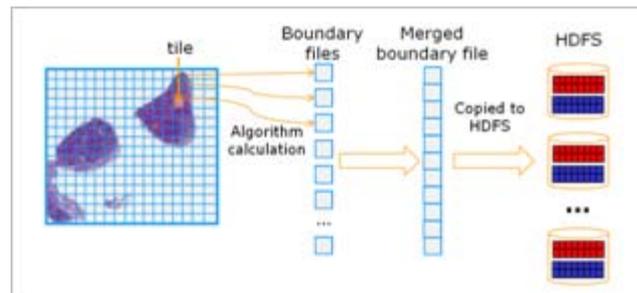


PAIS model

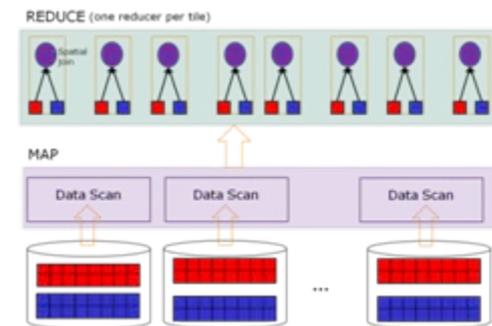


PAIS data management

Modeling and management of markup and annotation for querying and sharing through parallel RDBMS + spatial DBMS



HDFS data staging



MapReduce based queries

On the fly data processing for algorithm validation/algorithm sensitivity studies, or discovery of preliminary results

# PAIS Database and Analysis Pipeline



- Suite of analysis algorithms and pipelines that carry out the following tasks:
  1. segmentation of cells and nuclei;
  2. characterization of shape and texture features of segmented nuclei;
  3. storage of nuclei meta-data in relational database;
  4. mechanism supporting spatial queries for human-annotated nuclei;
  5. machine learning methods that integrate information from features to accomplish classification tasks.

# Pipeline for Whole Slide Feature Characterization



- $10^{10}$  pixels for each whole slide image
- 10 whole slide images per patient
- $10^8$  image features per whole slide image
- 10,000 brain tumor patients
- $10^{15}$  pixels
- $10^{13}$  features
- *Hundreds of algorithms*
- *Annotations and markups from dozens of humans*

# Integrative Research Design Pattern: Other Emory Examples



- Minority Health Genomics and Translational Research Bio-Repository Database (MH-GRID)
- ACTSI Cardiovascular, Diabetes, Brain Tumor Registry
- Early Hospital Readmission
- CFAR (Center for AIDS Research) HIV/Cancer Project
- Radiation Therapy and Quantitative Imaging
- Integrative Analysis of Text and Discrete Data Related to Smoking Cessation and Asthma
- Semantic Query and Analysis of Integrative Datasets in Renal Transplant Clinical Studies (CTOT-C)

# Thanks to:

- **In silico center team:** Dan Brat (Science PI), Tahsin Kurc, Ashish Sharma, Tony Pan, David Gutman, Jun Kong, Sharath Cholleti, Carlos Moreno, Chad Holder, Erwin Van Meir, Daniel Rubin, Tom Mikkelsen, Adam Flanders, Joel Saltz (Director)
- **caGrid Knowledge Center:** Joel Saltz, Mike Caliguiri, Steve Langella co-Directors; Tahsin Kurc, Himanshu Rathod Emory leads
- **caBIG In vivo imaging team:** Eliot Siegel, Paul Mulhern, Adam Flanders, David Channon, Daniel Rubin, Fred Prior, Larry Tarbox and many others
- **In vivo imaging Emory team:** Tony Pan, Ashish Sharma, Joel Saltz
- **Emory ATC Supplement team:** Tim Fox, Ashish Sharma, Tony Pan, Edi Schreibmann, Paul Pantalone
- **Digital Pathology R01:** Foran and Saltz; Jun Kong, Sharath Cholleti, Fusheng Wang, Tony Pan, Tahsin Kurc, Ashish Sharma, David Gutman (Emory), Wenjin Chen, Vicky Chu, Jun Hu, Lin Yang, David J. Foran (Rutgers)
- **NIH/in silico TCGA Imaging Group:** Scott Hwang, Bob Clifford, Erich Huang, Dima Hammoud, Manal Jilwan, Prashant Raghavan, Max Wintermark, David Gutman, Carlos Moreno, Lee Cooper, John Freymann, Justin Kirby, Arun Krishnan, Seena Dehkharghani, Carl Jaffe
- **ACTSI Biomedical Informatics Program:** Marc Overcash, Tim Morris, Tahsin Kurc, Alexander Quarshie, Circe Tsui, Adam Davis, Sharon Mason, Andrew Post, Alfredo Tirado-Ramos
- **NSF Scientific Workflow Collaboration:** Vijay Kumar, Yolanda Gil, Mary Hall, Ewa Deelman, Tahsin Kurc, P. Sadayappan, Gaurang Mehta, Karan Vahi





**caBIG**<sup>™</sup> *cancer Biomedical  
Informatics Grid*<sup>™</sup>

an initiative of the National Cancer Institute

Thanks!