Data Fusion and & 21st Century Diagnostics

Fused Diagnostics
Integrated Diagnostics
Advanced Analytics
Computational Diagnostics
Personalized Predictive Modeling

John E. Tomaszewski, MD

Professor and Chair Department of Pathology & Anatomical Science,
University at Buffalo, State University of New York

Learning Objectives

- Recognize that informative data can exist in multiple modes and at multiple levels of scale.
- Understand that diagnostics in the future will likely depend on the integration or fusion of data from multiple sources

Disclosures

- NIH R01CA1365235-01 “Software to allow for multimode, multiscale fused data for pathology and radiology”
- “ Histology based computer assisted diagnostics (HistoCAD) for automated grading of prostate cancer”. Sponsored Research Agreement and licensed software to Bioimagene, closed 4/1/11
- Scientific Advisory Board member, IbRiS Inc., no financial relationships.

The Collaborative Group

Center for Computational Imaging and Personalized Diagnostics

Anant Madabhushi, PhD

Postdocs:
- James Monaco, PhD
- Gaoyu Xiao, PhD
- Jun Xu, PhD

Graduate Students:
- Jonathan Chappelow
- Scott Doyle
- Satish Viswanath
- Pallavi Tiwari
- George Lee
- Shannon Agner
- Ajay Basavanhally
- Rob Toth
- Andrew Janowczyk
- Akshay Shridar
- Najeeb Chowdhury
- Sahir Ali
- Rachel Sparks

Undergraduate Students
- Pratik Patel
- Elaine Yu

Clinical Collaborators
- Michael D. Feldman, MD, PhD
- John E Tomaszewski, MD
- Mark Rosen, MD, PhD
- Shridar Ganesan, MD, PhD
- Nicholas Bloch, MD
- Steven Master, PhD, MD
- John Kurhanewicz, PhD

Case Western Reserve University, Est. 1826

University of Pennsylvania School of Medicine, Department of Pathology & Laboratory Medicine

UCSF University of California San Francisco

Rutgers University

UB University at Buffalo, The State University of New York

“Personalized Medicine”

Form 1:

Targeted Therapies Directed at Genetic Variations-Pharmacogenomics

Form 2:

Cellular Engineering and Vaccine Therapy

Form 3:

Fused Diagnostics

Personalized Medicine Form 1

Pharmacogenomics:

Genomewide RNA expression, DNA copy-number and sequence analyses, and microRNA and proteomic profiling — have the potential to allow individualized selection of treatment as determined by the characteristics of the patient and the tumor.

The NEW ENGLAND JOURNAL of MEDICINE

EDITORIAL

Personalized Medicine and Inhibition of *EGFR* Signaling in Lung Cancer

Adi F. Gazdar, M.D.

EGFR Mutations and Response to TK Inhibitors

- Mutations in the tyrosine kinase domain of EGFR result in ligand independent gene activation
  - Deletion of a conserved sequence in exon 19 of EGFR
  - Point mutation in exon 21 of EGFR (L858R)

- The nonrandomized European study by Rosellet al. shows the feasibility of large-scale screening for *EGFR* mutations in patients with advanced non–small-cell lung cancer for selection for *erlotinib* therapy.

EGFR Mutations and Response to TK Inhibitors

- Many confounders
  - Role of copy number
  - Not all mutations with same effect
  - Activating mutations in KRAS confer resistance
  - Response vs. survival

Personalized Medicine Form 2

Cellular and Vaccine Engineering:

- Using autologous cells as a target for cellular engineering and reconstitution of patient with these modified cells for a desired specific therapeutic effect.
- “It doesn’t get more personal than using your own cells to cure your disease”

Personalized Rx for HIV

Carl June et al , U of Pa and Sangamo Bioscience

- 1% of Caucasians are resistant to HIV infection
- Mutation in CCR5 which encodes receptor on CD4 T cells needed for HIV to infect these cells
- Zinc finger targeting endonucleases to the CCR5 gene with a yield of CCR5-ve cells that are resistant to HIV infection
- Clinical trials underway

The Digital Transformation of Glass

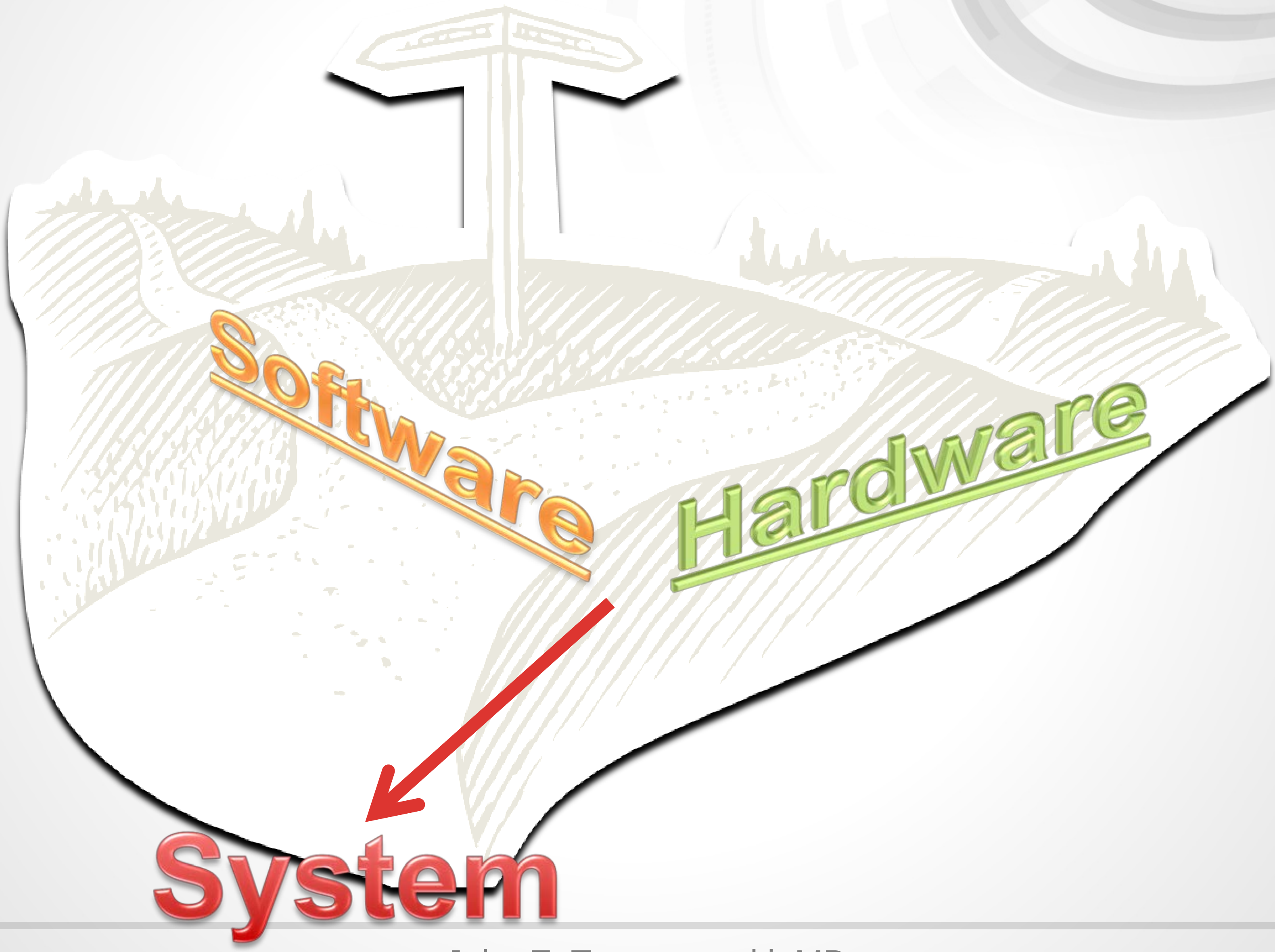
One Element of Fused Diagnostics

- – **Virtual slides**
- – **ROI finding algorithms**
- – **Automated assistive analysis**
- – **Computational biology**

Developments in microscope-based pathology

1700's	Microscopy
1850's	Hematoxylin + Cell theory of disease and cancer (Virchow)
1870's	Eosin
1890's	Formaldehyde
1940's	Immunohistochemistry (direct fluorescence)
1970's	In-situ hybridization (radioactive DNA)
1990's	The computer

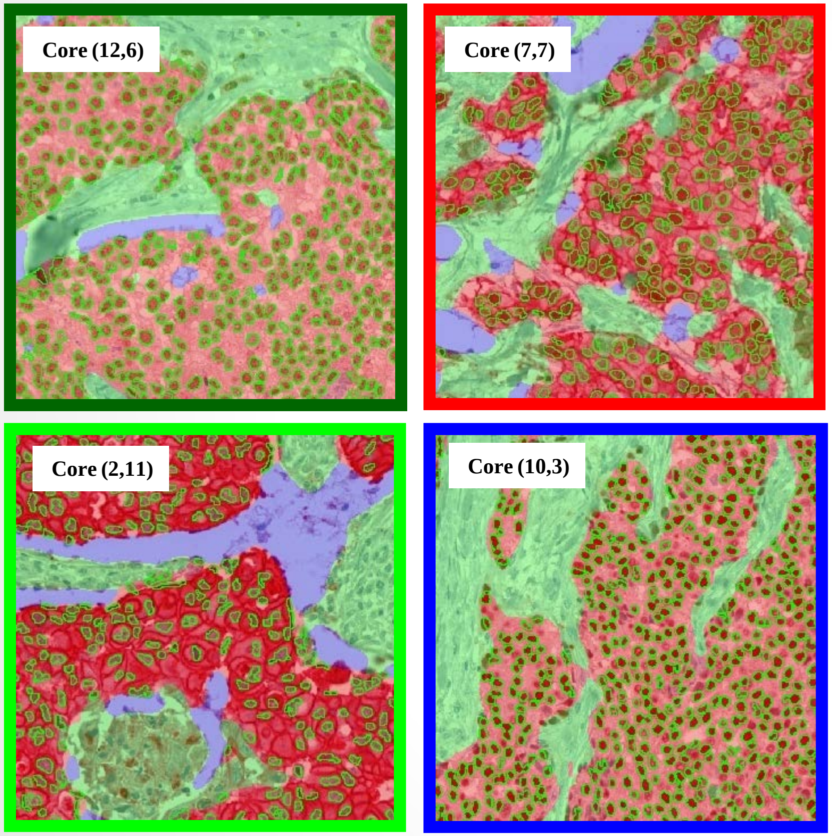
Quantitative Histocytometry

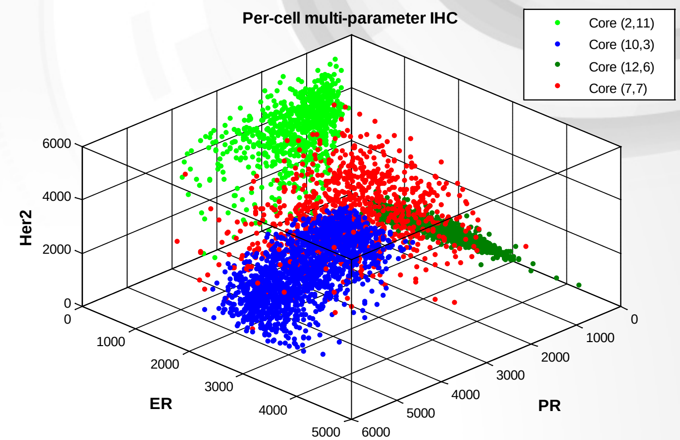


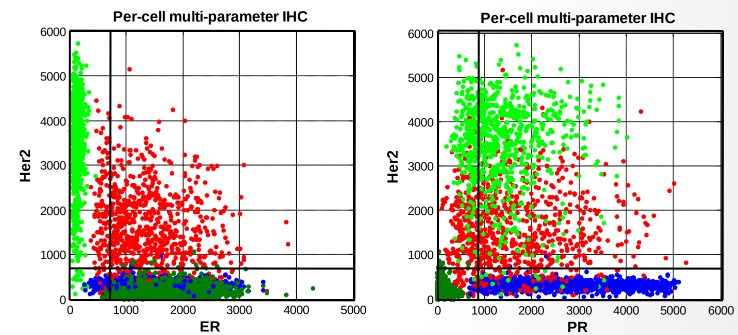
John E. Tomaszewski, MD

Multi-Parameter IHC Demonstration with ER, PR, and Her2

Classify tissue with machine learning algorithms, segment cellular compartments, extract spectrally unmixed signals from associated compartments, and export per-cell protein expression.







ER vs Her2:

0%	0%
2%	98%

11%	75%
1%	13%

98%	0%
2%	0%

0%	0%
8%	92%

PR vs Her2:

1%	0%
99%	0%

16%	69%
3%	12%

2%	84%
0%	14%

0%	0%
0%	100%

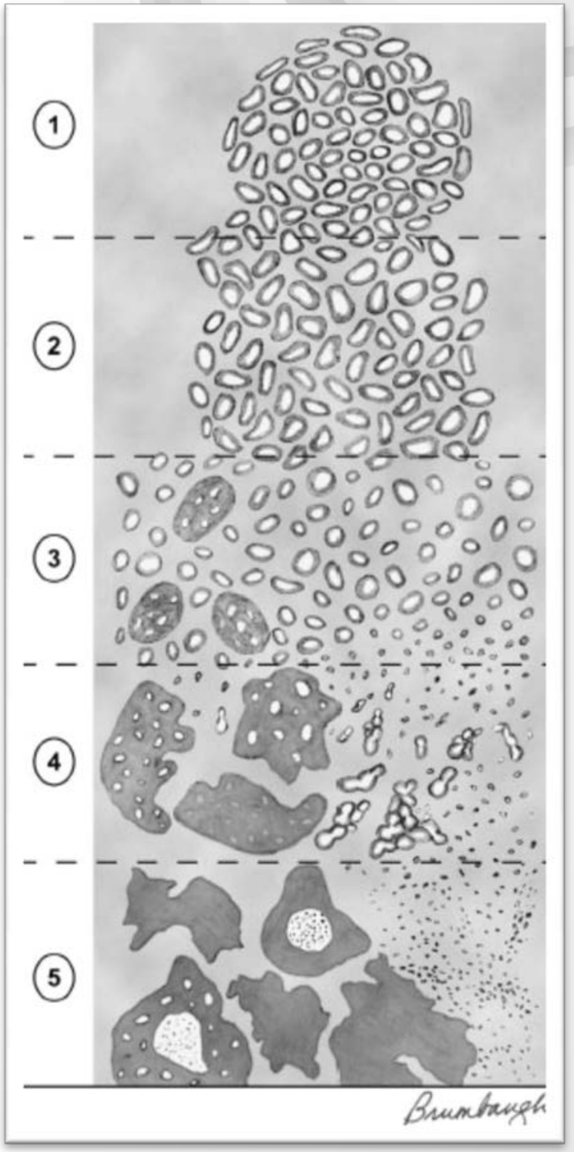
3-D scatter plots show multi-parameter data extracted from individual cells. Note distinct clusters. Cell data can be assessed with cluster or quadrant analysis, commonly used in flow cytometry for phenotyping (double negative, single positive, double positive , etc.)

Quantitative Analysis
Machine Vision and Learning
Computer Assisted Diagnosis (CAD)

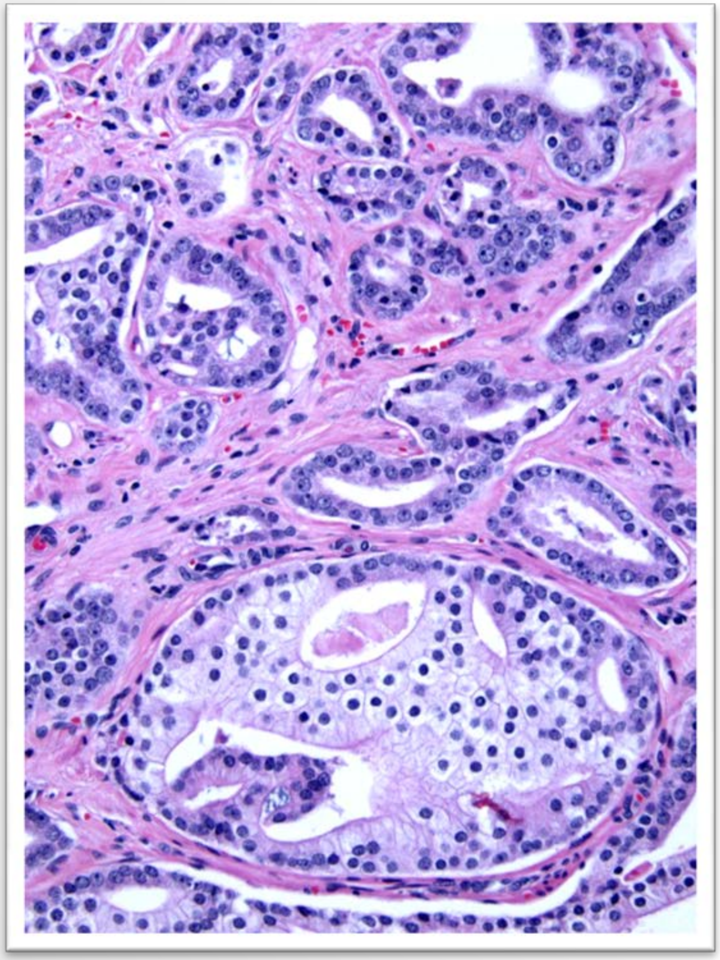
John E. Tomaszewski, MD

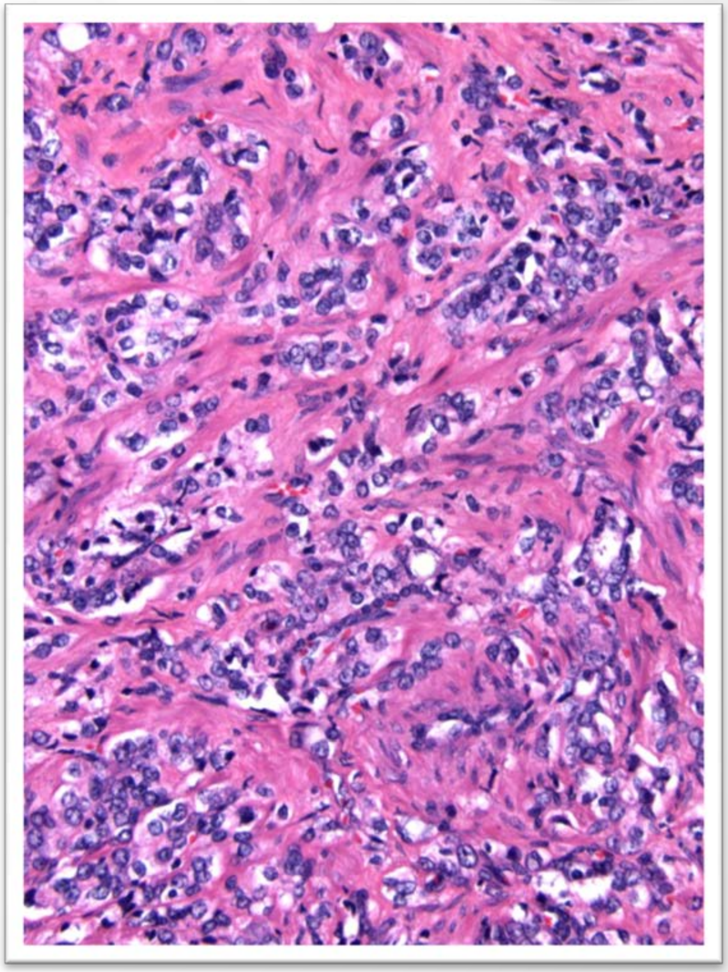
A Challenge
Identifying Significant Prostate Cancer

Gleason Grading Identifies Five Primary Hisopathological Patterns of Gland Growth (Micron Resolution Imaging)

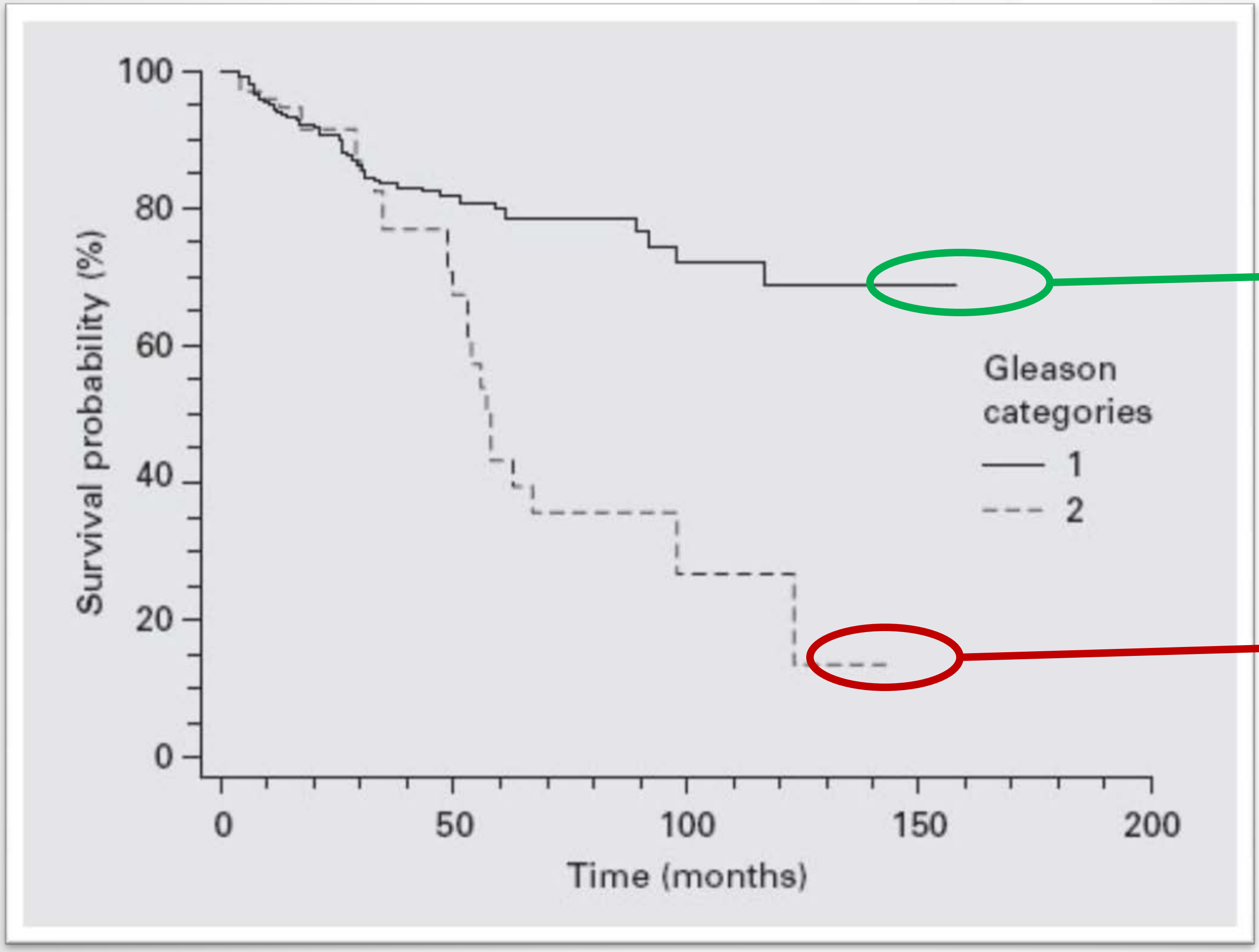


Pattern 3 vs. Pattern 4/5





Gleason Sum is Strong Predictor of Clinical Progression



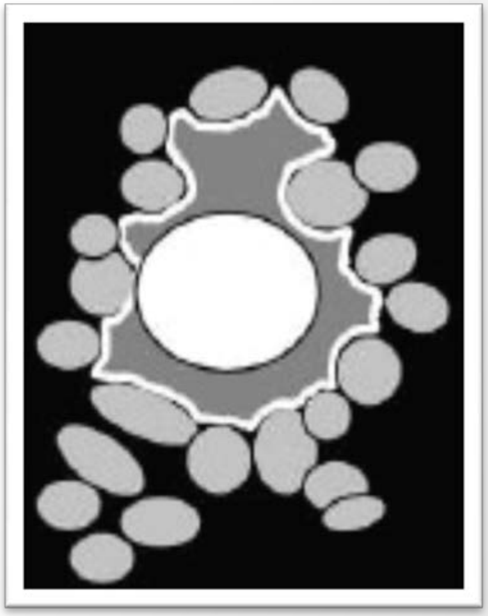
Gleason Sum 6

Gleason Sum >7

Pinto et al. Urol Int 7:202-208, 2006

Can CAD be Used to Find Gleason Pattern 4 CAP ?

Find the Neoplastic Glands



A predicate structure:
Computer Assisted Gland Segmentation Focused on interior gland lumens and centroid of each gland

Segmenting Glands

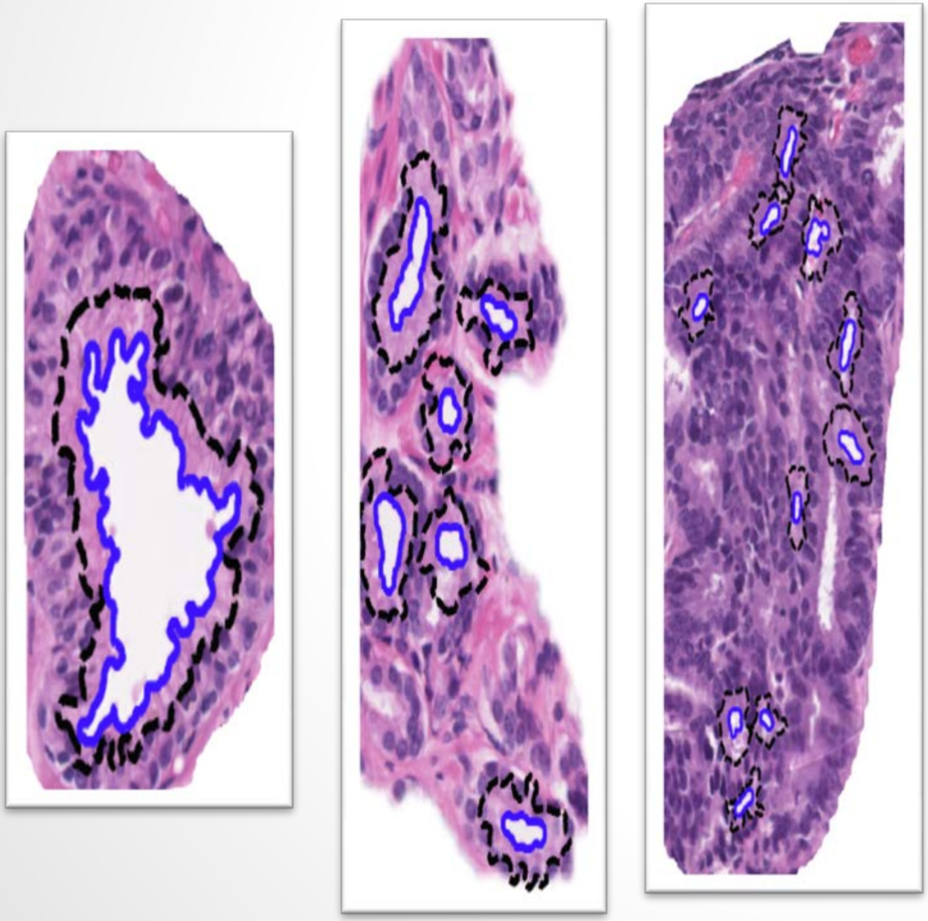


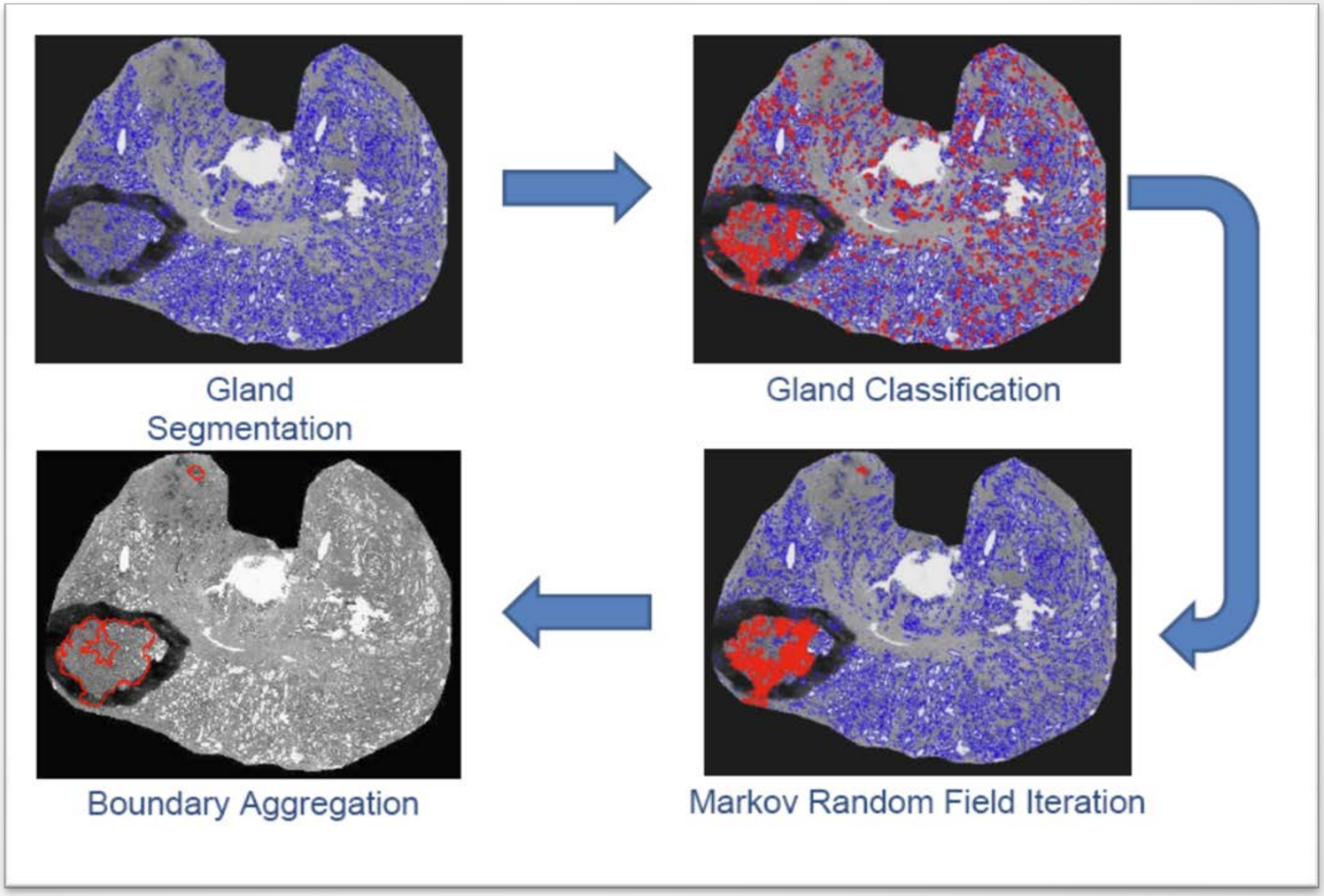
Figure 2: Results of the automatic segmentation algorithm (blue contours – lumen boundary, black contours -- inner boundary of the nuclei of the epithelial cells surrounding the gland). Shown from left to right are example images of benign epithelium, intermediate-, and high grade cancer.

Classification of Individual Glands

- Leverages two features
  - Gland size , malignant glands are “microacini”
  - The tendency for proximate glands to share the same class
  - This spatial dependency is modeled using a Markov prior
    - Describes the local inter-site dependencies of a random process
    - “Cancer is clonal”

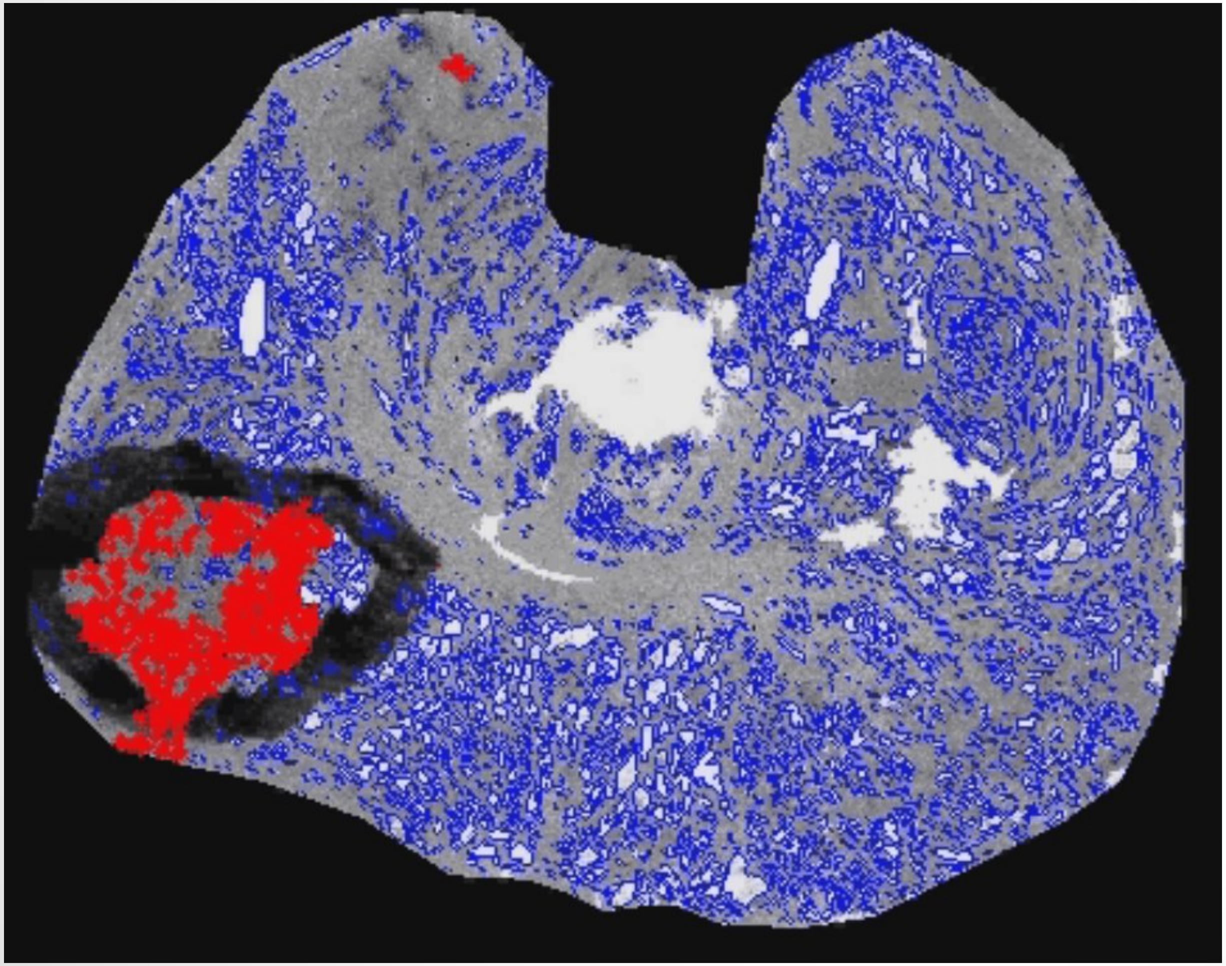
Region finding in whole mount

Automated region growing gland segmentation algorithm



Monaco J., et al, "Detection of prostate cancer from whole-mount histology images using Markov random fields", MICCAI, NY, 2008

Overview of CaP Detection Algorithm



Gland Segmentation

⇩

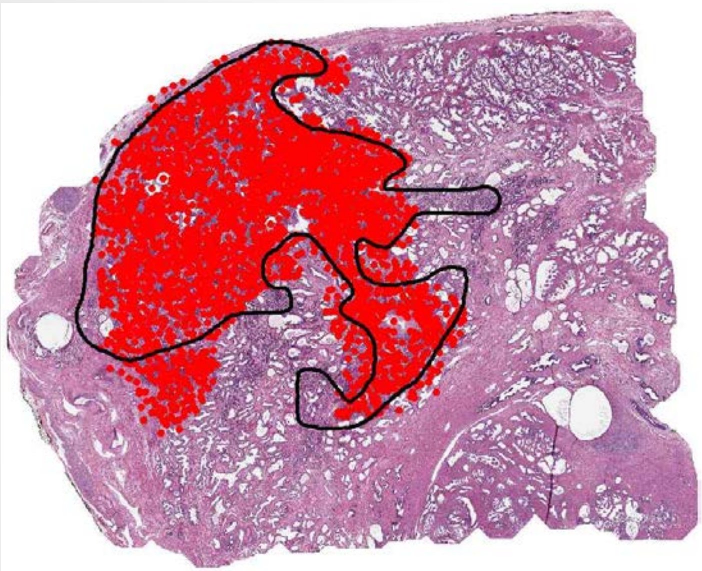
Gland Classification

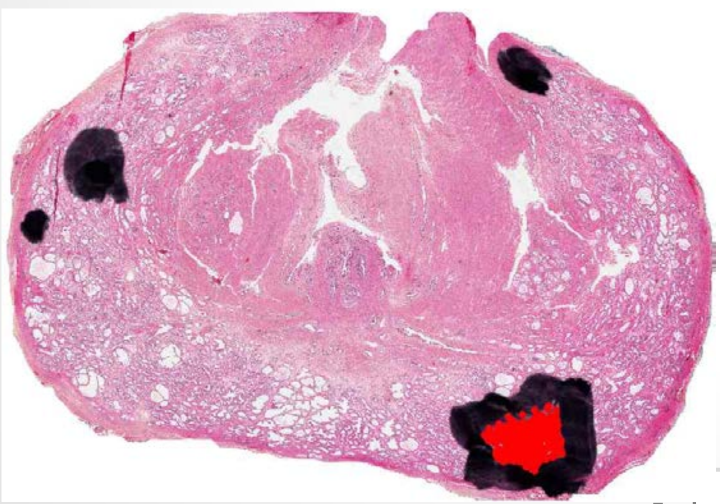
⇩

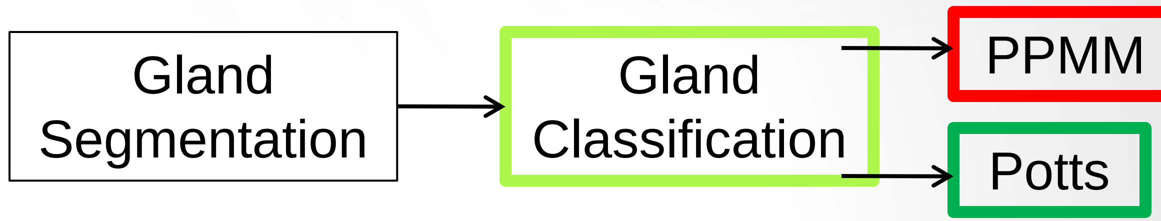
Markov Random Field Iteration

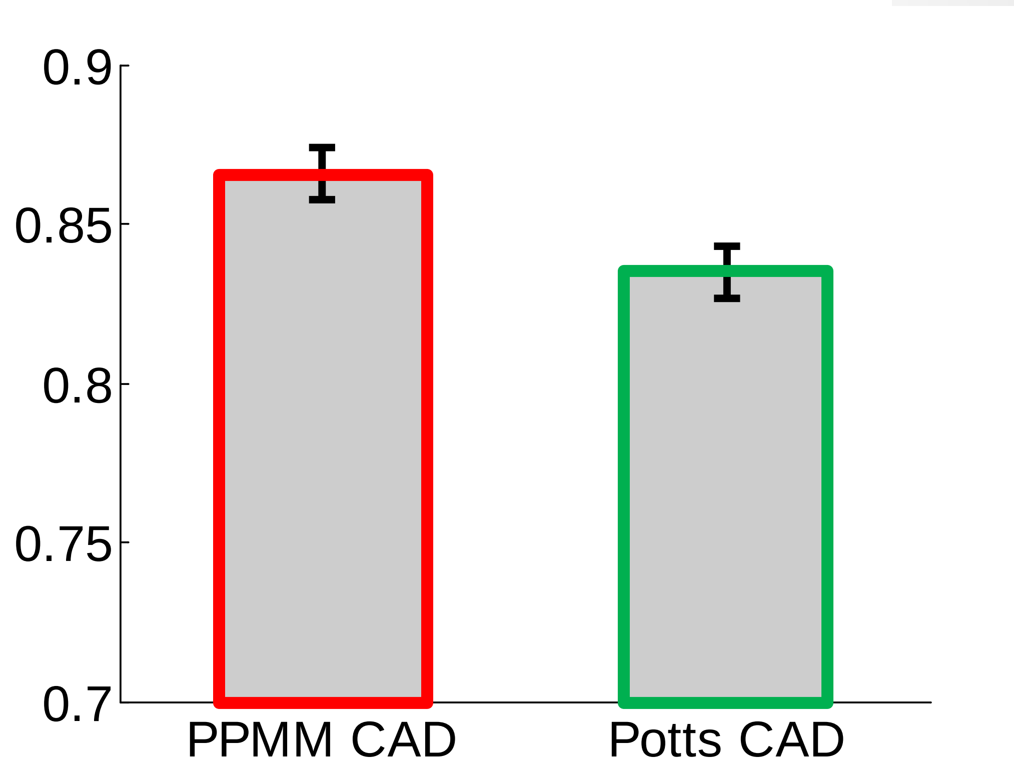
Monaco J et al. Medical Image Analysis 14:617, 2010

CAD Results









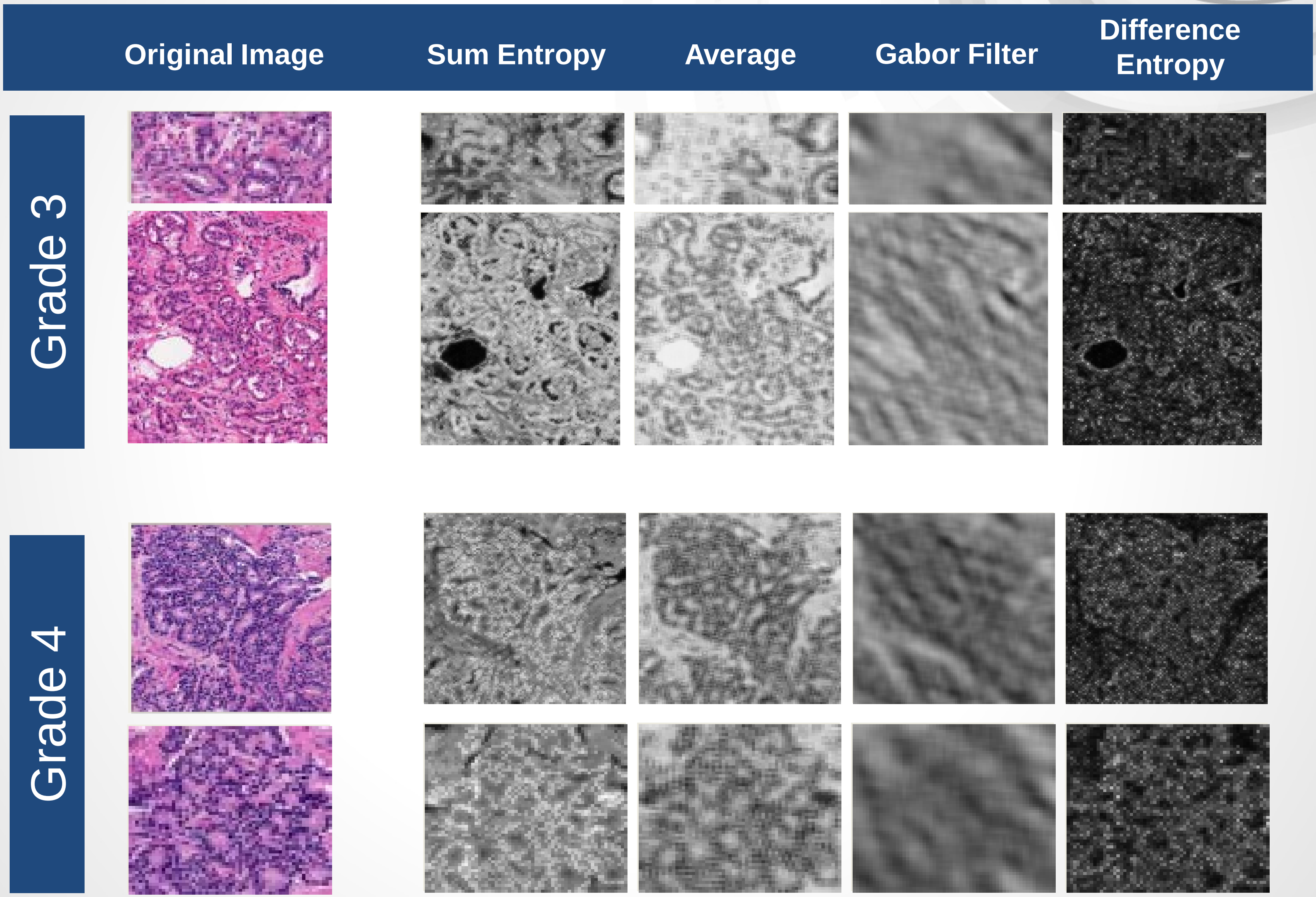
Monaco J et al. Medical Image Analysis 14:617, 2010

Can CAD be Used to Find Gleason Pattern 4 CAP ?

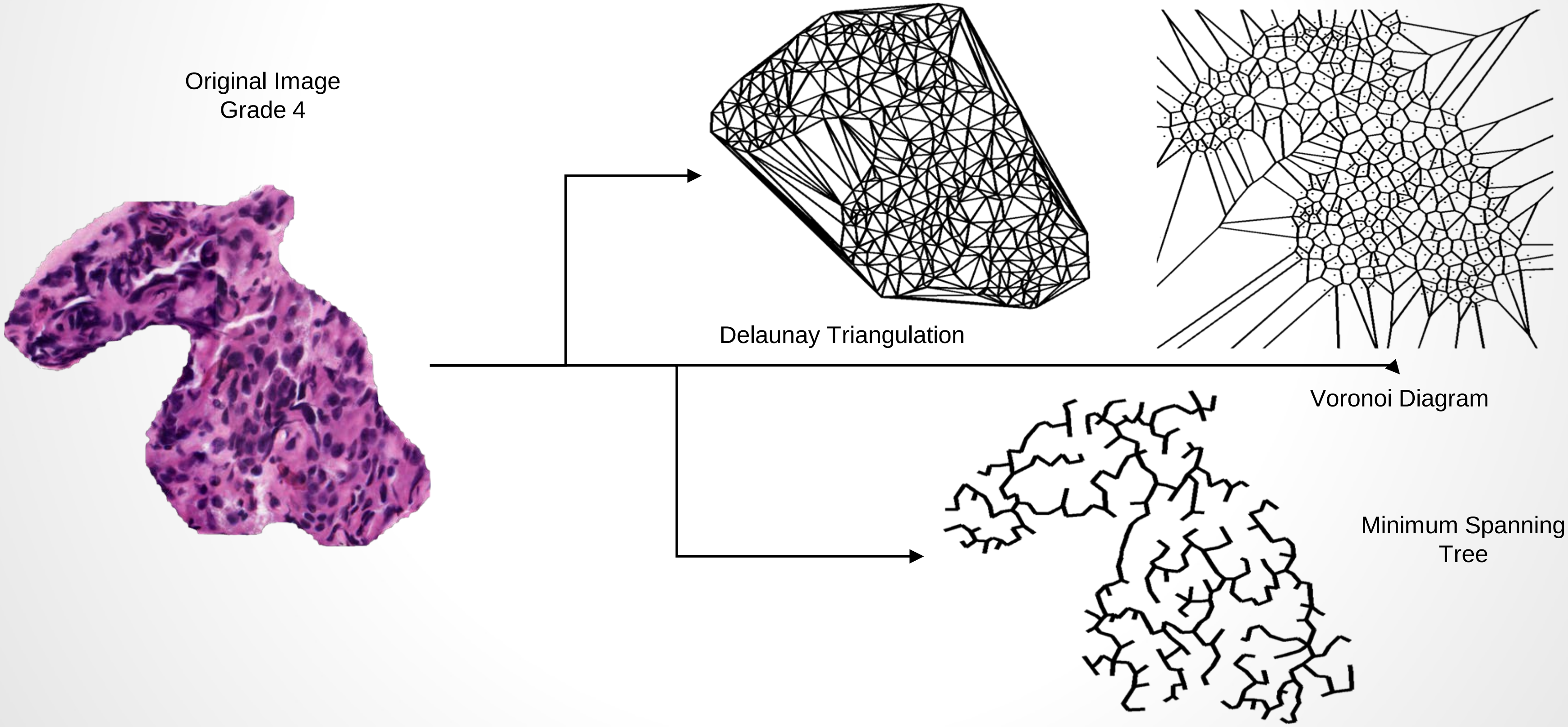
Establish Informative Quantitative Features

Feature Type	Number of features	Gleason Feature
Nuclear arrangement	25	Nuclear proliferation and infiltration
Graph Features	24	None
Gland Morphology and Architecture	44	Gland differentiation and organization
Texture (Statistical, Haralick, Gabor)	483	None

Doyle S, Hwang M, Shah K, Madabhushi A, Feldman M, Tomaszewski J. IEEE, 2007


Original Image
Sum Entropy
Average
Gabor Filter
Difference Entropy
Grade 3
Grade 4

Graph Embedding



An Image Analysis Approach to CAP Grading

Texture Features Examining Pixels

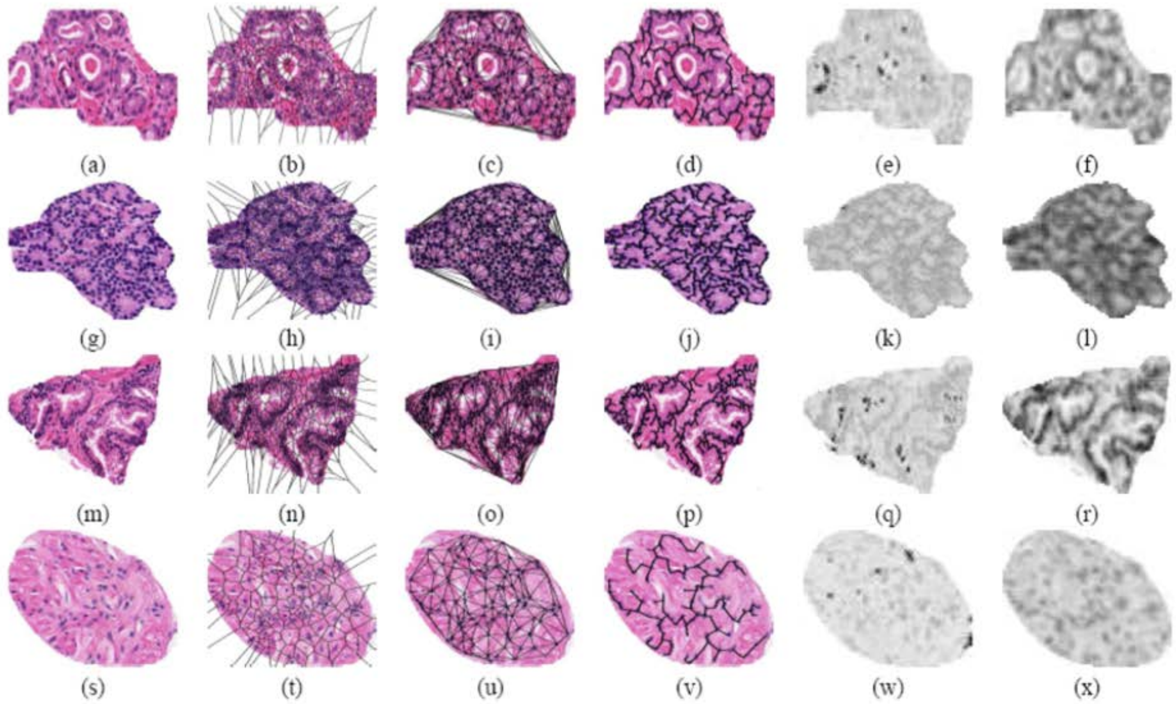


Fig. 3. Comparison of ((a)-(f)) Gleason grade 3 tissue, ((g)-(l)) grade 4 tissue, ((m)-(r)) benign epithelium, and ((s)-(x)) benign stroma. Superimposed on ((a), (g), (m), (s)) the original images are ((b), (h), (n), (t)) the Voronoi diagram, ((c), (i), (o), (u)) the Delaunay triangulation, ((d), (j), (p), (v)) the minimum spanning trees, ((e), (k), (q), (w)) pixel entropy texture feature, and ((f), (l), (r), (x)) Gabor filter ($s = 3$, $\theta = \frac{5\pi}{8}$).

Geometric Features Interrogating Nuclei or Glands

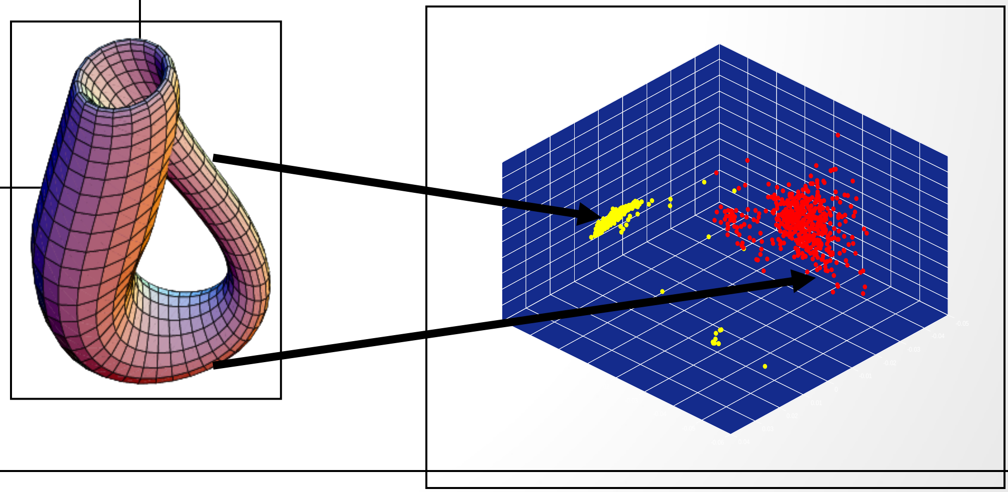
“The Curse of Multidimensionality”

- In Large Multidimensional Space One Can Always Find Multiple Solutions to a Two Class Problem (CAP or Not CAP)
- In Order to Avoid This Multiplicity of Solutions One Must Reduce the Dimensions of the Data Used for Classification
- Manifold Learning Methods Reduce the Dimensionality of a Data Set from *N* Dimensions to *M* Dimensions where *N*>>>*M*

Dimensionality Reduction Mapping Functions

Non-linear data dimensionality reduction, unlike linear methods (PCA) -- object proximity is preserved.

- Locally Linear Embedding
- Isometric Mapping
- Spectral Clustering



Investigating manifold learning schemes for analyzing various very high dimensional biomedical data.

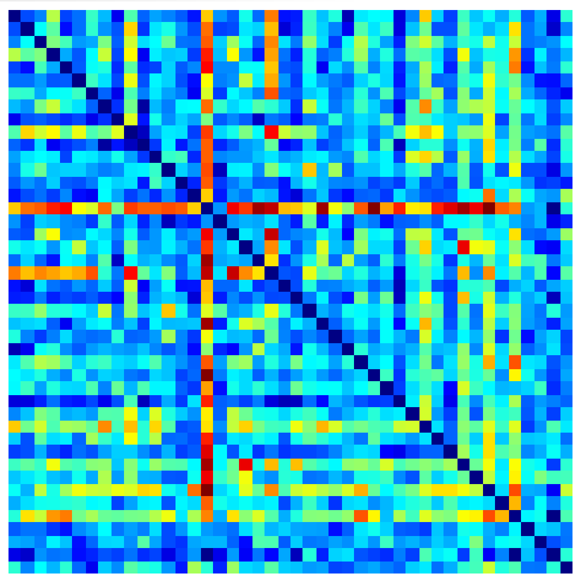
Dimensionality Reduction Methods

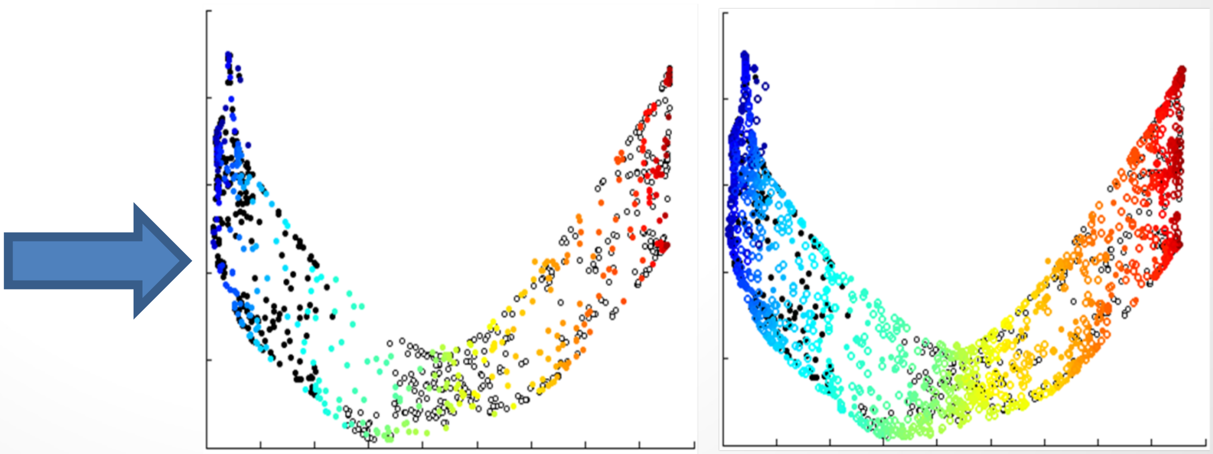
- Linear
  - **Principle component analysis**

- Non-Linear
  - **Graph Embedding**
    - **Preserves geodesic distances between images**
    - **Transforms high dimensionality dataset into low-dimensionality space**

Non-linear mapping functions: Graph Embedding

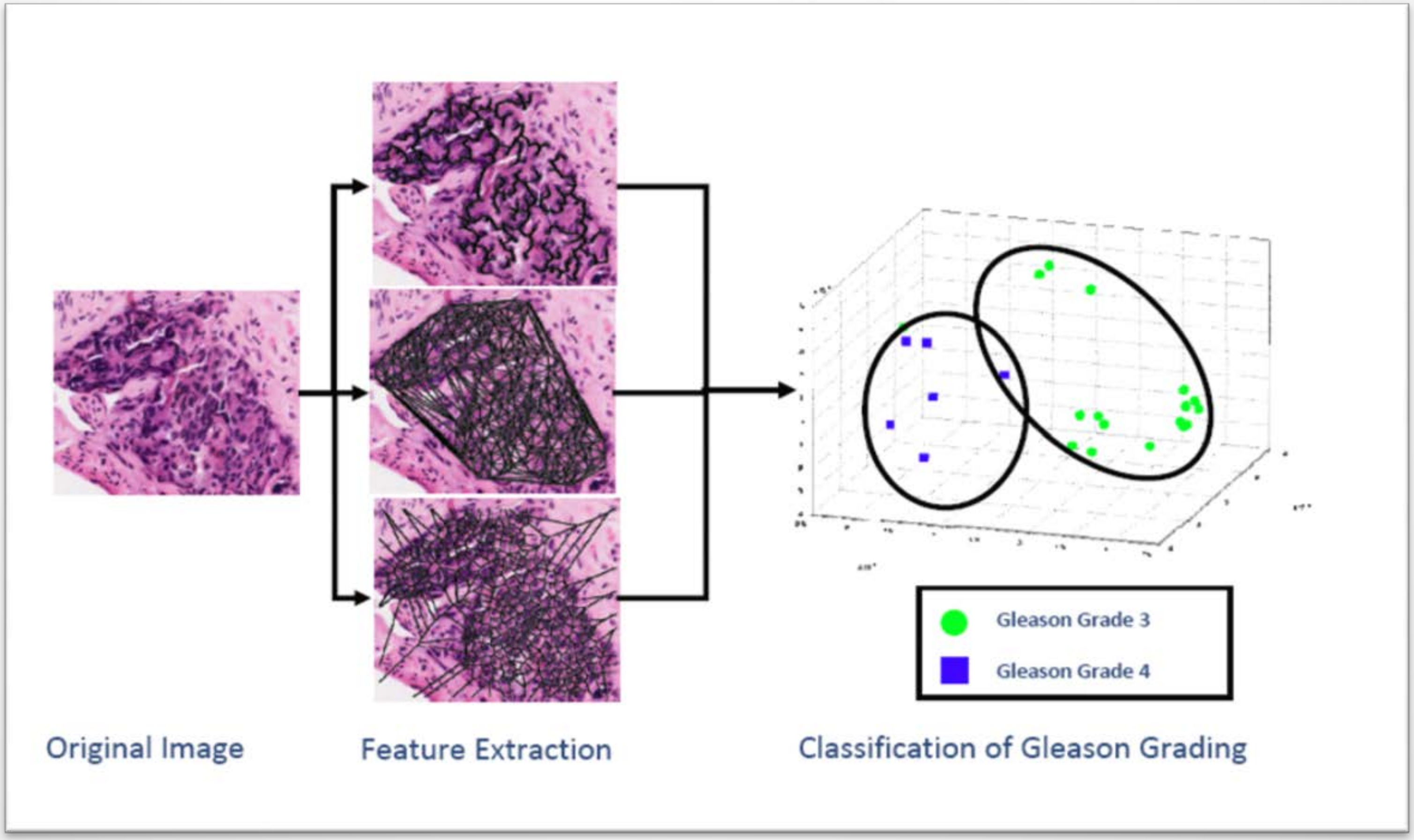
- GE allows for projecting data into reduced dimensional space.
- SVM classifier applied in reduced dimensional space.



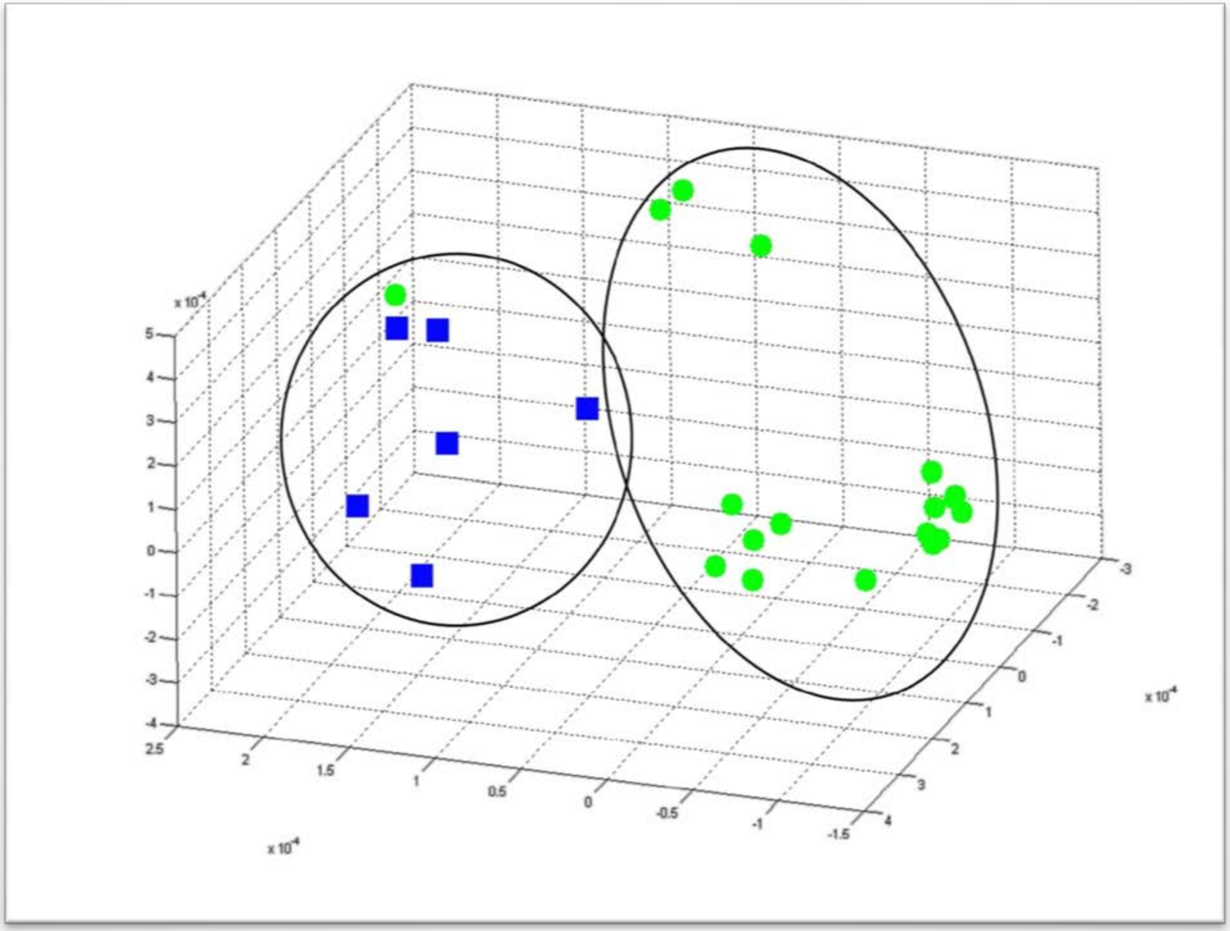


Shi et al, PAMI 1998

Grading Prostate Carcinoma



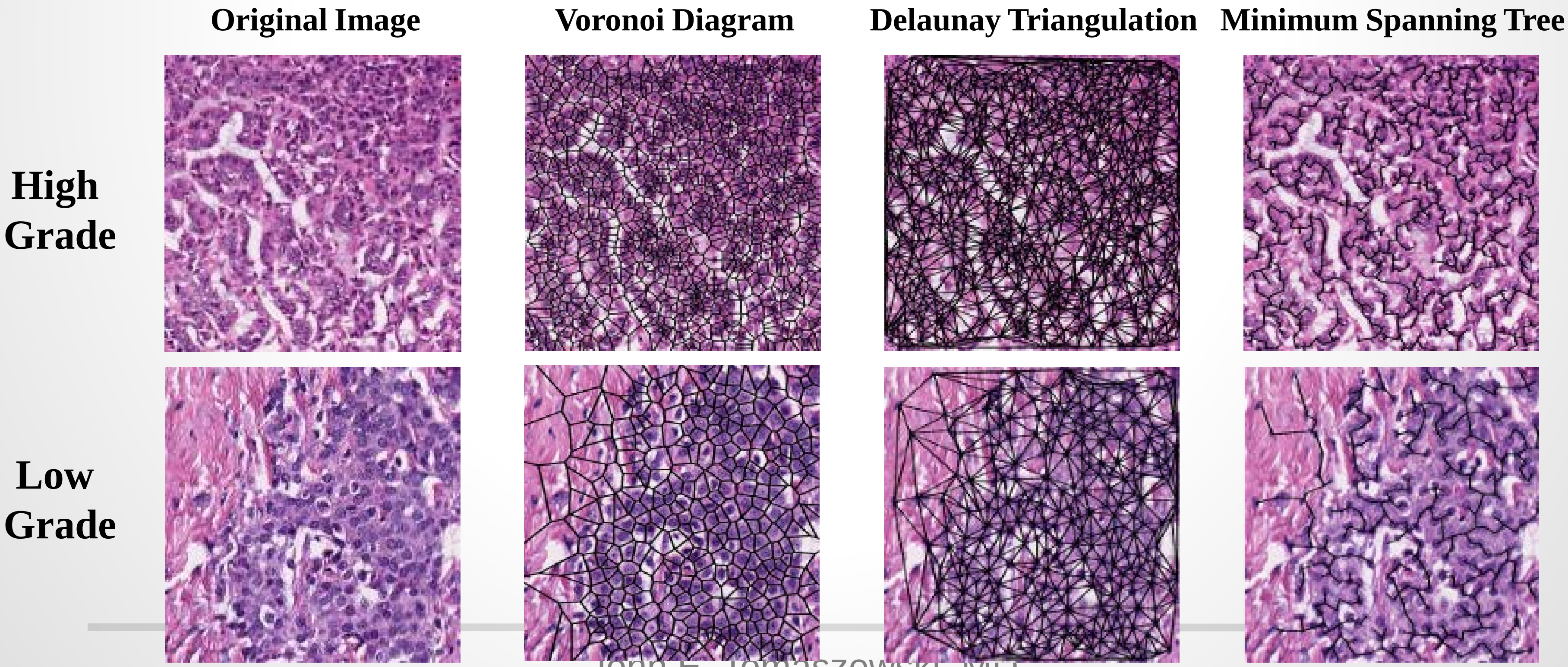
Graph Embedding Analysis of CAD Data with Dimensionality Reduction and Separation of Gleason Patterns 3 and 4



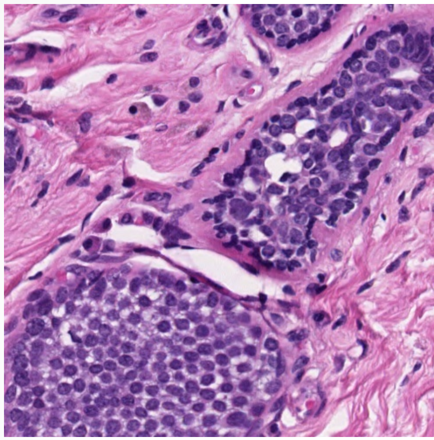
John E. Tomaszewski, MD

Breast Cancer Tissue Grading

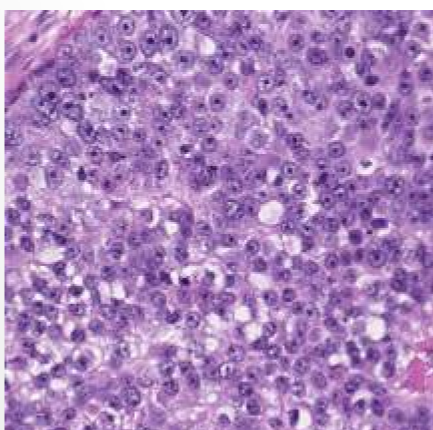
- Aim: Quantify tissue structures to distinguish Bloom-Richardson grades of breast cancer



Visualizing Breast Graph Features



Low Grade



High Grade

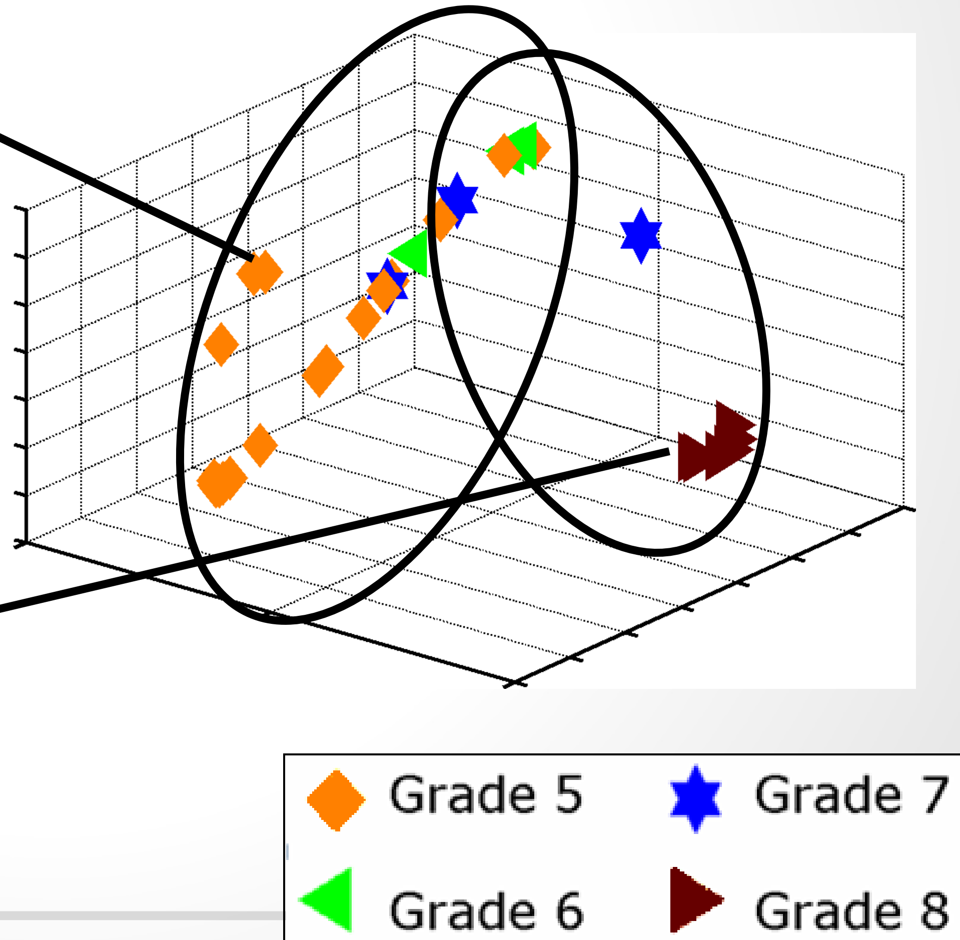
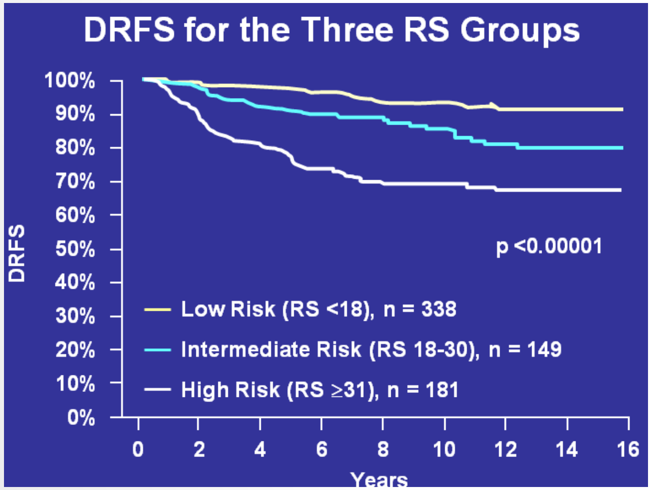


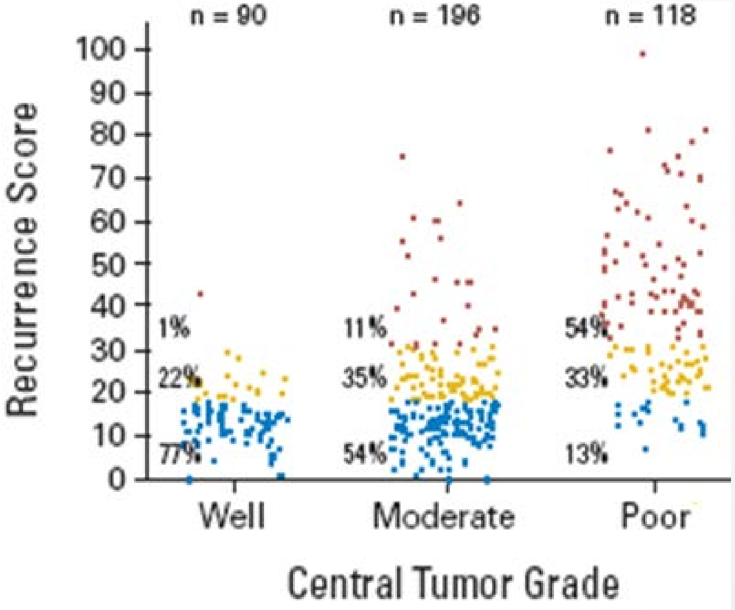
Image-based Risk Score (IbRiS)

- **Aim:** Create an Image-based Risk Score to predict recurrence of ER+ breast cancer



Recurrence via gene assay

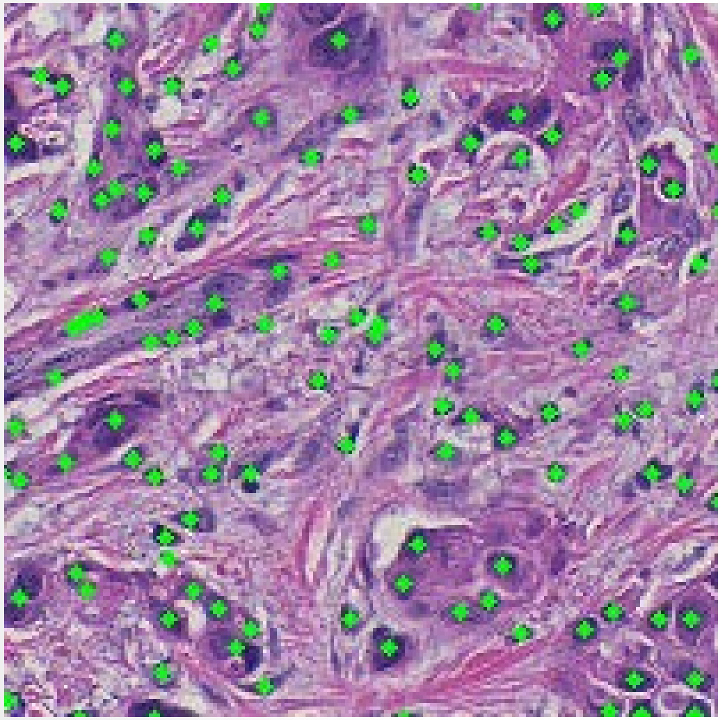
Paik et al., A Multigene Assay to Predict Recurrence of Tamoxifen-Treated, Node-Negative Breast Cancer. N Engl J Med 2004 351: 2817-2826



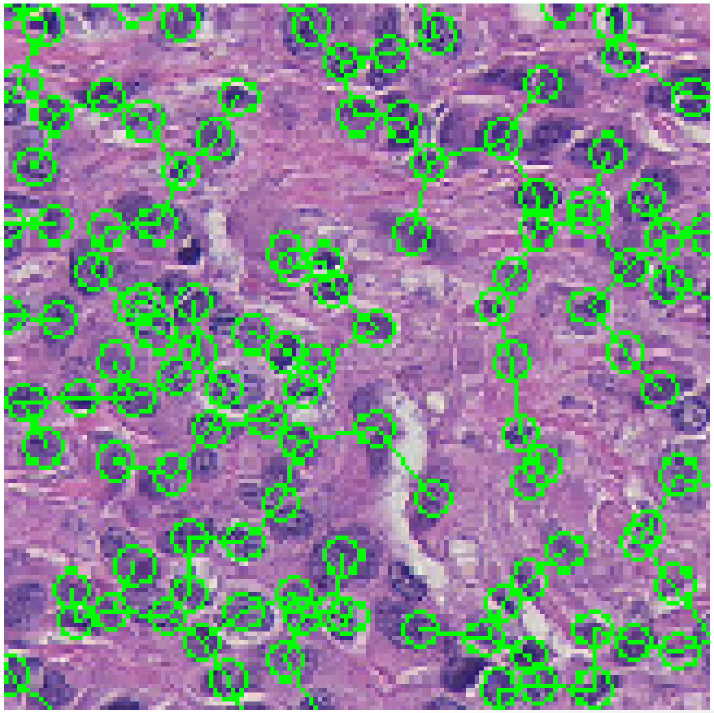
Recurrence via histology grade

Badve, Sunil S., et al., "Estrogen- and Progesterone-Receptor Status in ECOG 2197: Comparison of Immunohistochemistry by Local and Central Laboratories and Quantitative Reverse Transcription Polymerase Chain Reaction by Central Laboratory," J Clin Oncol 2008 26: 2473-2481

Components of IbRiS

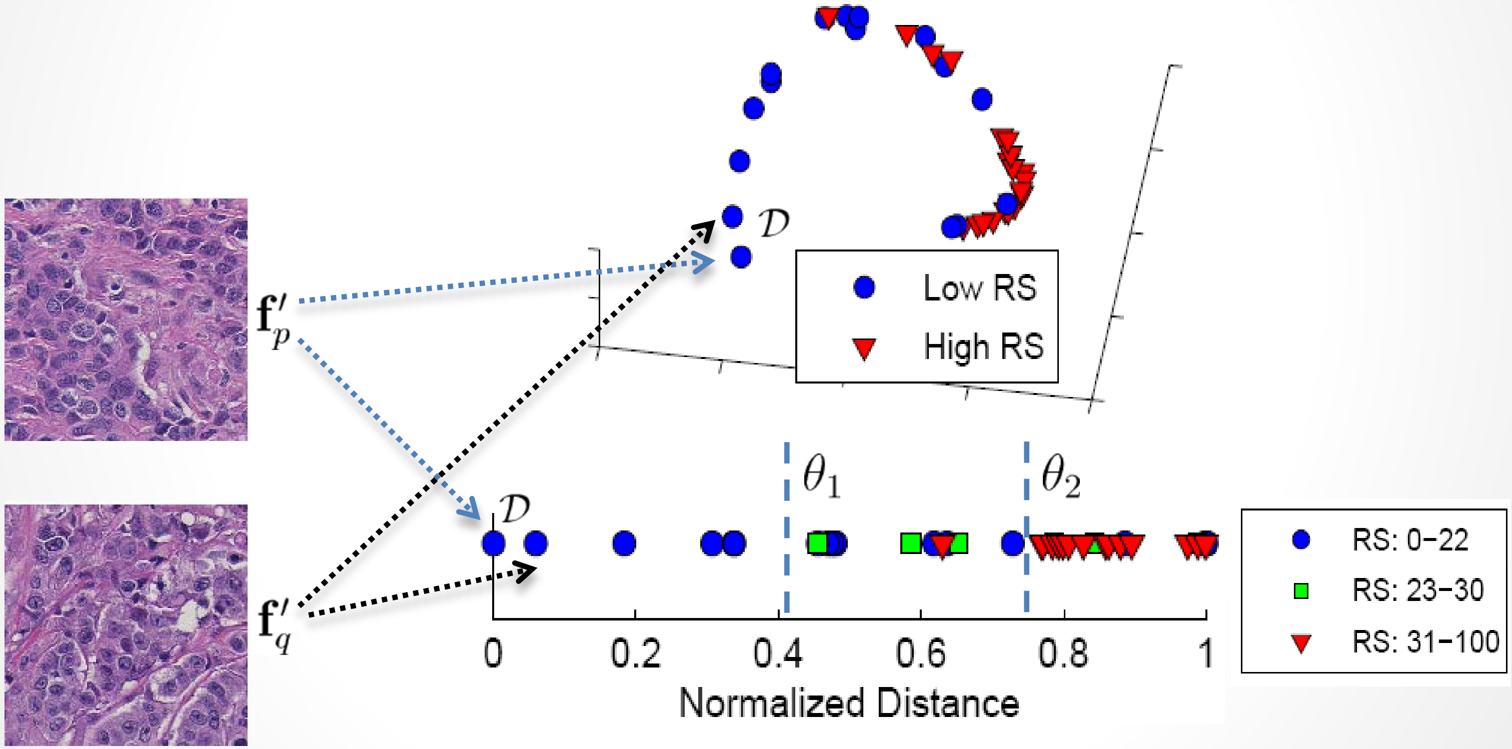


Automated Nuclei Detection



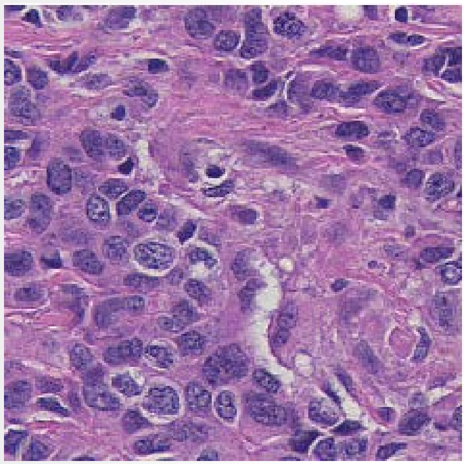
Graph Features Calculated

Obtaining IbRiS from Graph Features

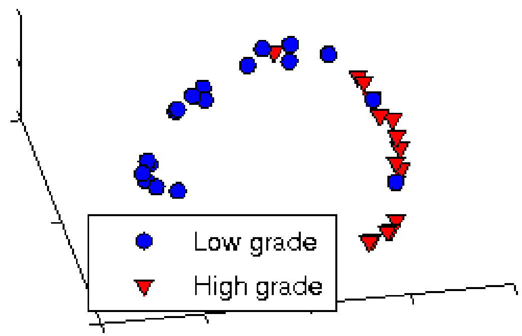


Basavanhally, A, Xu, J, Madabhushi, A, Ganesan, S, "Computer-aided prognosis of ER+ breast cancer histopathology and correlating survival outcome with Oncotype DX assay," Biomedical Imaging: From Nano to Macro, pp.851-854, 2009.

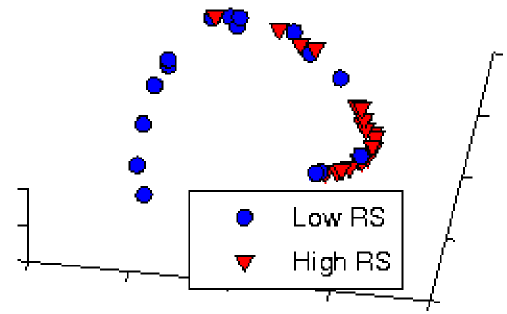
Comparing IbRiS to Genomic Assay

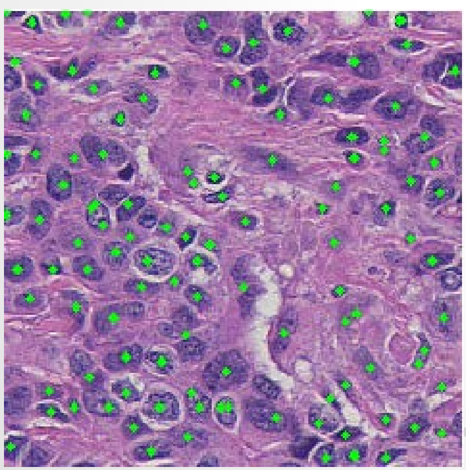


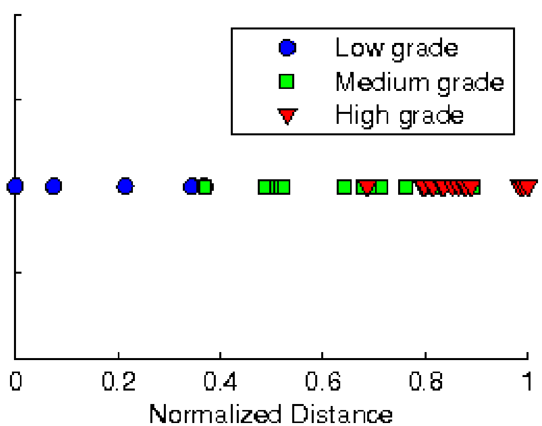
Breast cancer Grade based discrimination

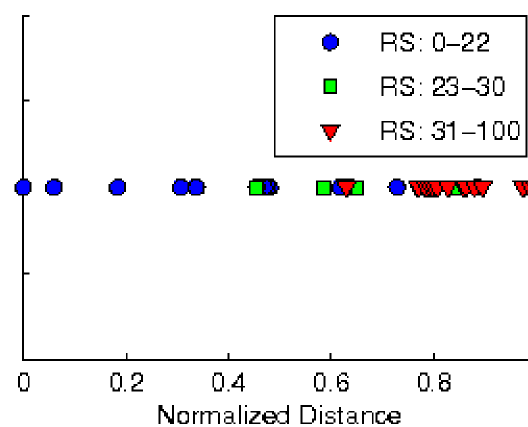


Oncotype Dx









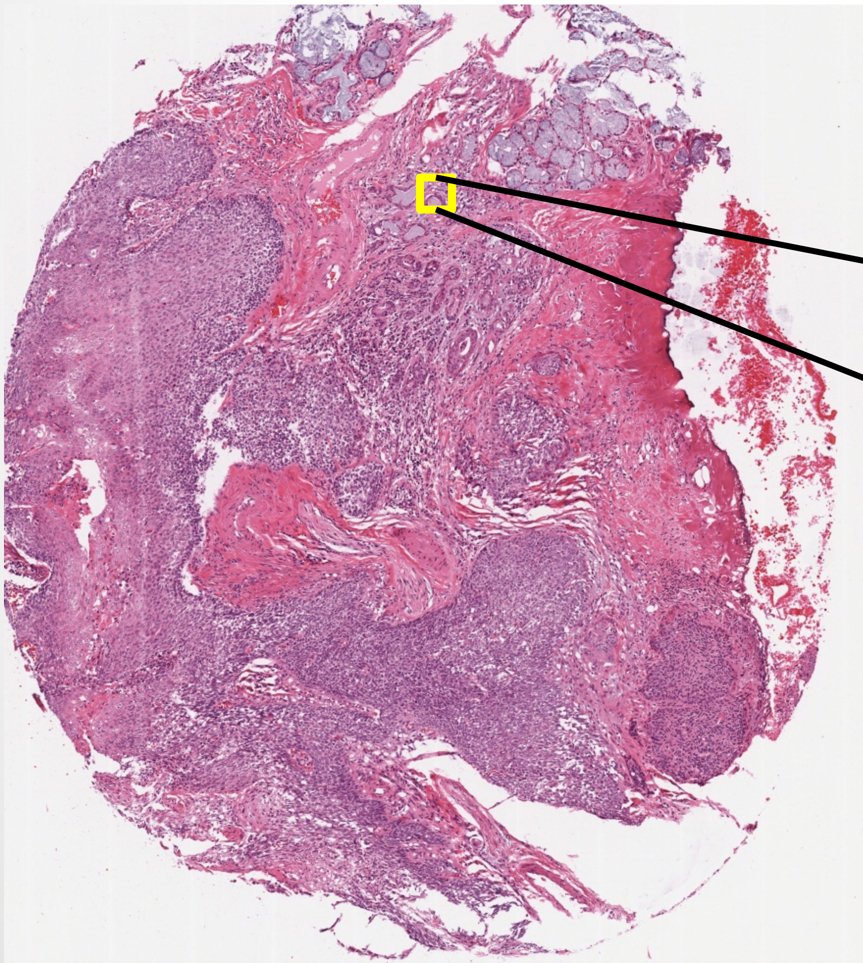
ANOTHER USE CASE FOR QUANTITATIVE HISTOMORPHOMETRY:

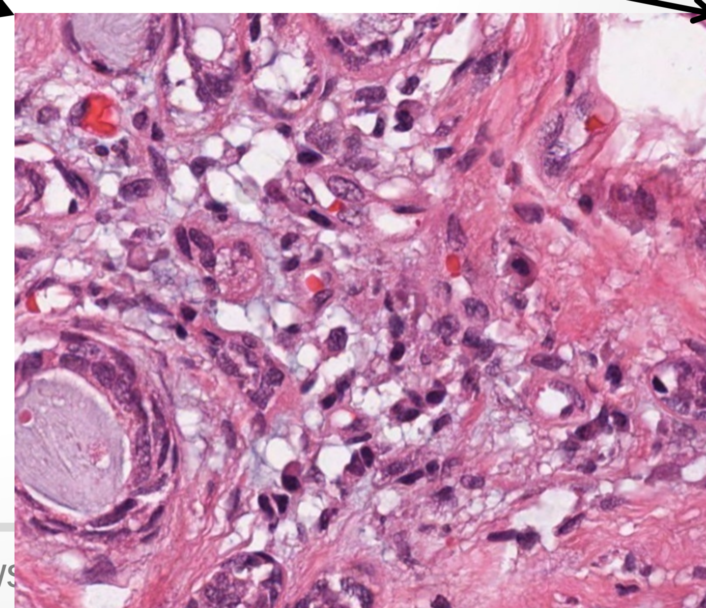
PREDICTING SURVIVAL IN HPV OROPHARYNGEAL TUMORS

John E. Tomaszewski, MD

Oropharyngeal Dataset

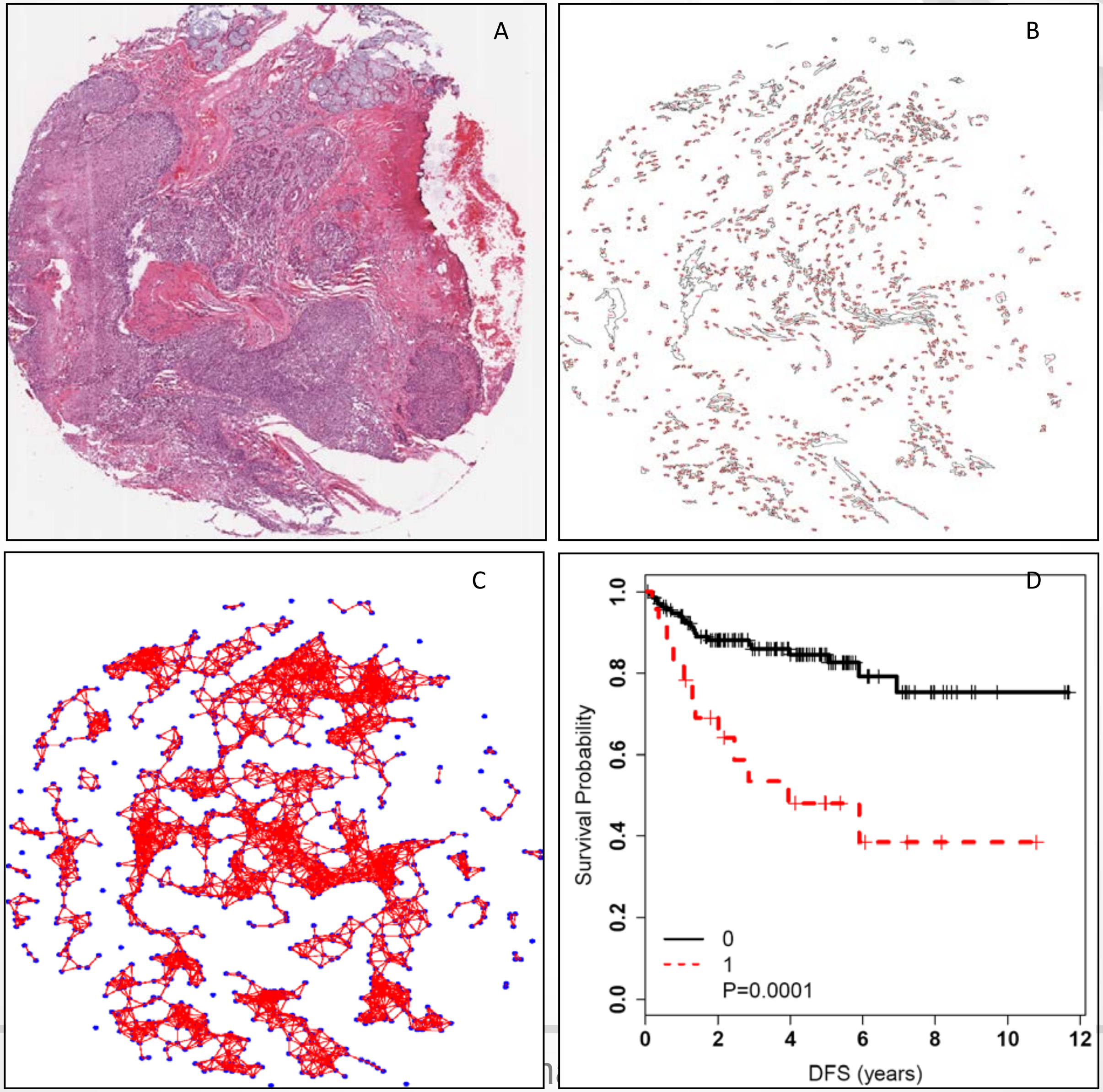
- HUGE TMAs (each around 11000 x 13000)
- p16+ Oropharyngeal SCC TMAs
- 26 Progressors, 124 Non progressors





John E. Tomaszews

CELL CLUSTER GRAPHS

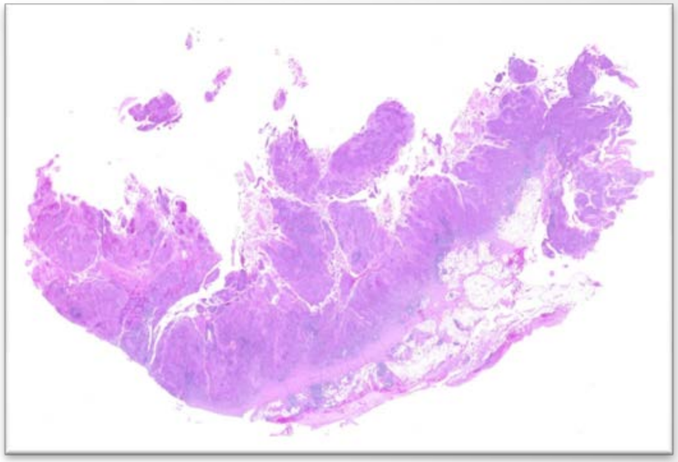


How to Meet Diagnostic Challenges of the Future: An Opinion

- In Finance , the Axiom is
  “ Follow the Money
- In Medicine …………
  “Follow the Data”

Characteristics of Data Used in the Diagnostics of the *Future*

- Extremely Large and Quantitative Data Sets
- Require Dimensionality Reduction
- Machine Learning
- Scaled Data
- Different Types of Data Converge – “Fused Diagnostics”



Path Image Data 2 GB

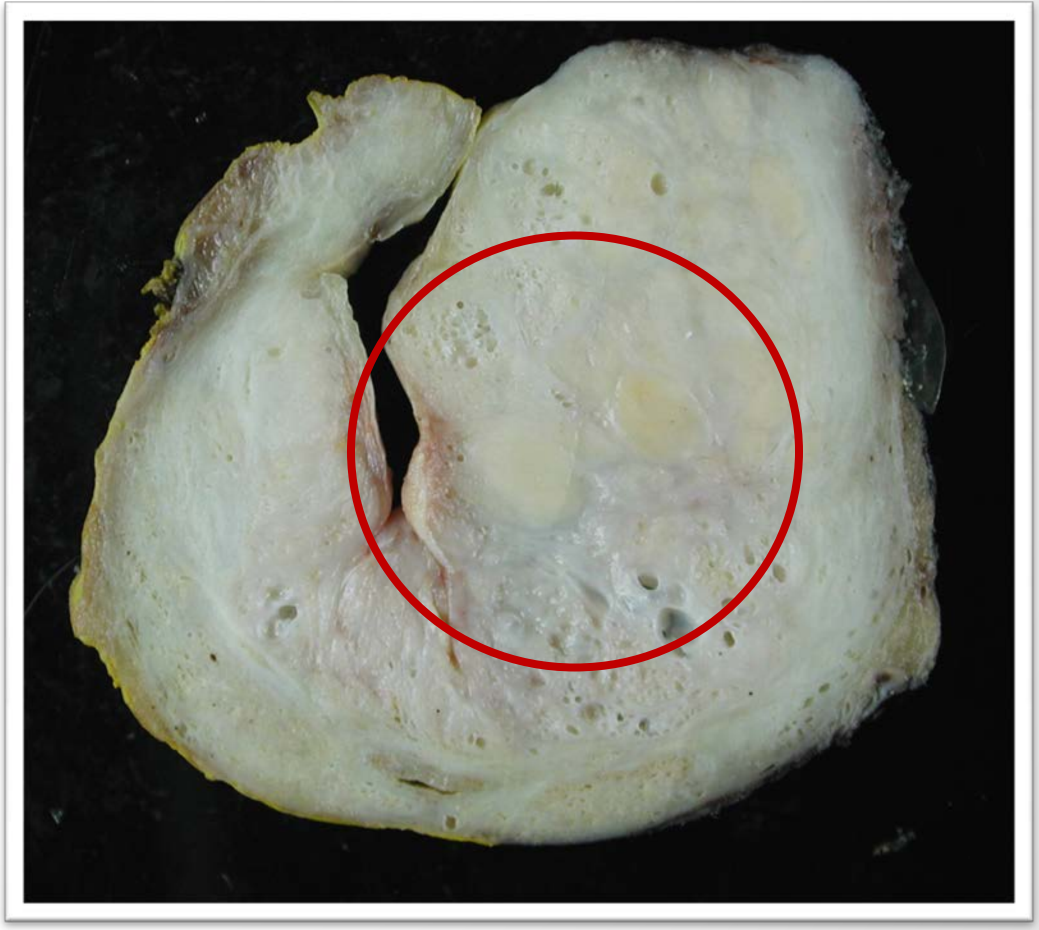
Microarray Data30K = 7.5 MB

Data is “Scaled”

- There is Informative Data at Every Resolution of Examination
- Informative Data at Each Level of Resolution can be Mapped to the Other Levels.
- The Data at Each Level of Resolution has Unique Attributes

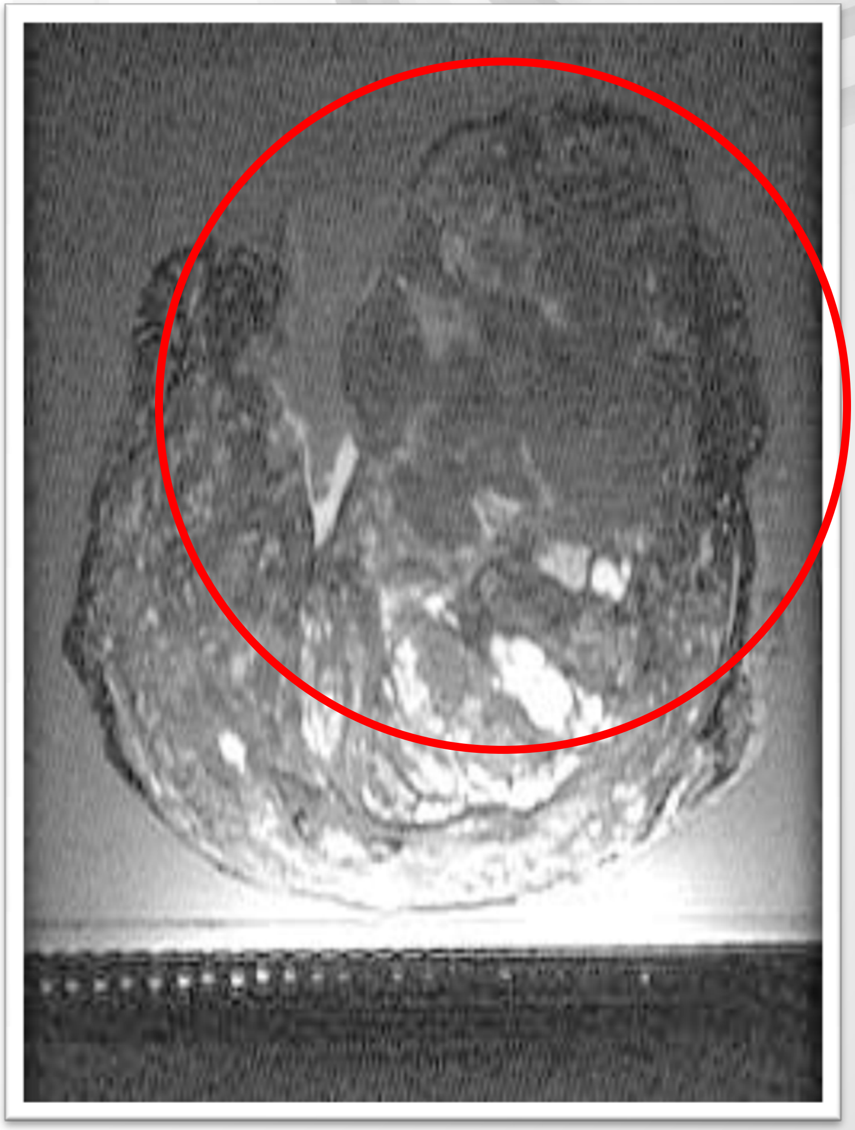
OPTICAL *1x MAGNIFICATION* RESOLUTION= 10^{-3}m

TEXTURE & COLOR VARIGATION SUSPICIOUS for CAP



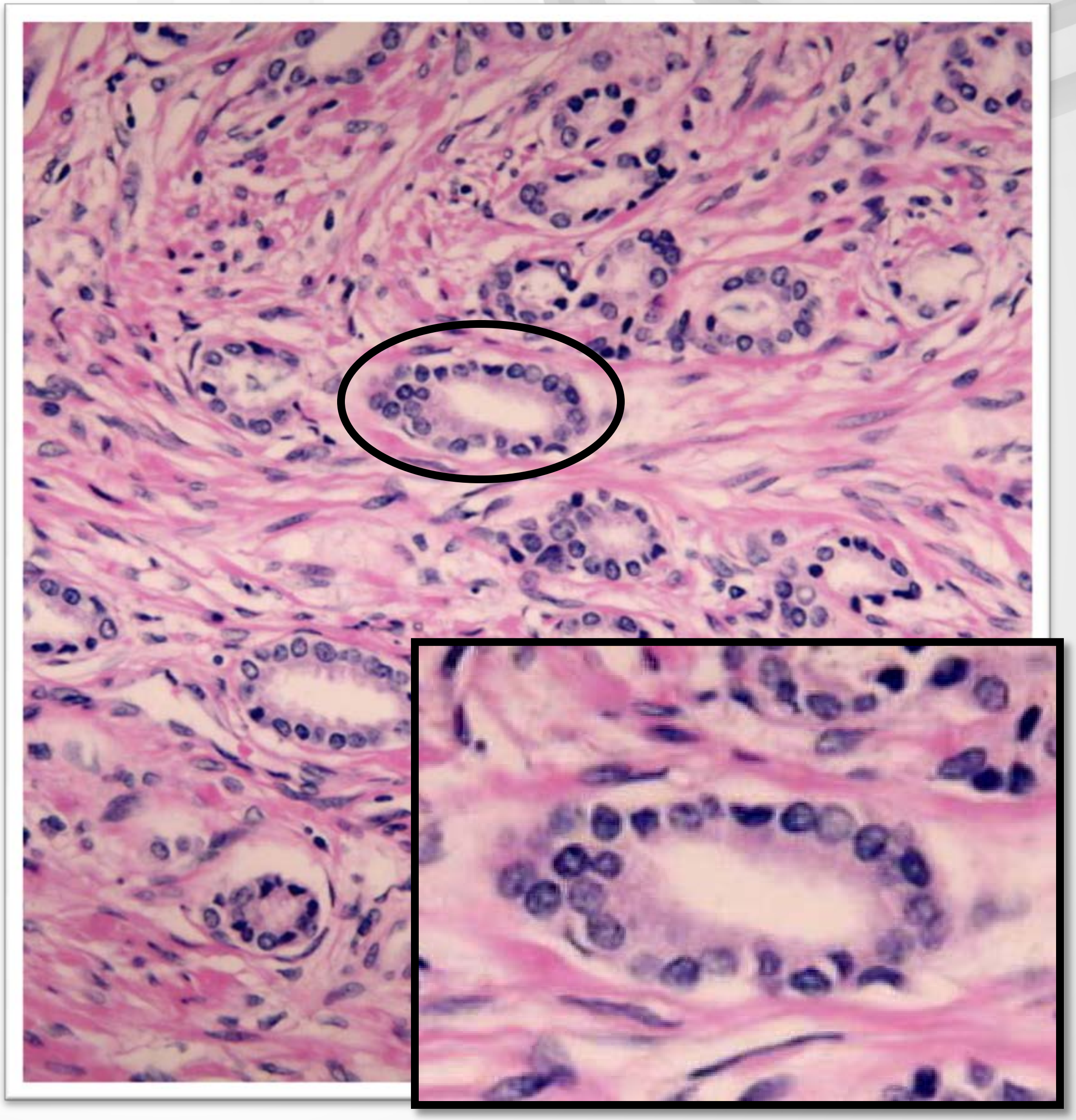
4T MR
T2
EX-VIVO
1x MAGNIFICATION
RESOLUTION=10^{-4}m

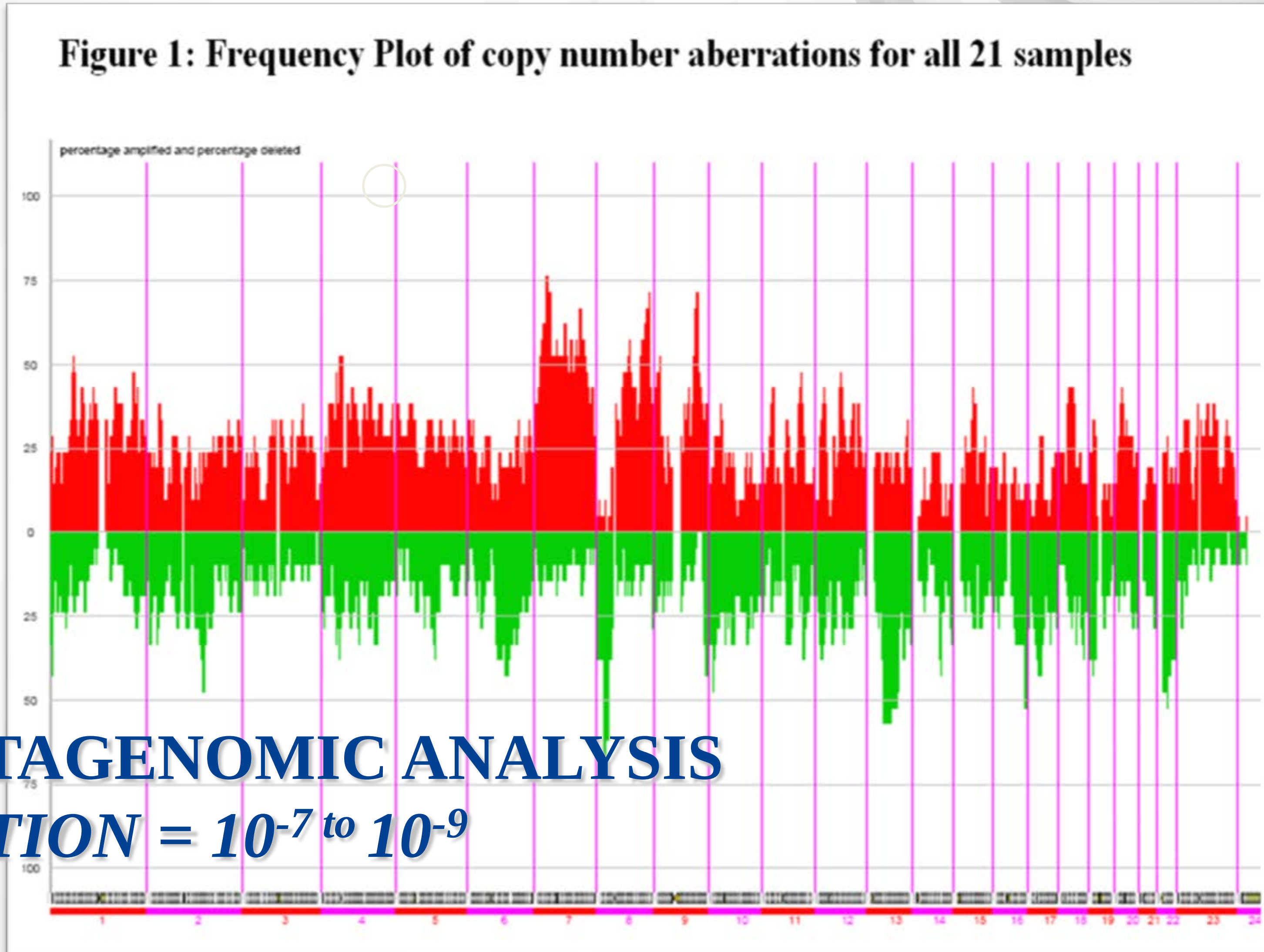
HYPODENSE
AREA INTERRUPTING
NORMAL CURVILINEAR
GLAND ARCHITECTURE
DIAGNOSTIC of CAP
(mostly)



OPTICAL
200X MAGNIFICATION
RESOLUTION=10^{-7}m

HAPHAZZARD PATTERN of INFILTRATING MICROACINI DIAGNOSTIC of CAP GLEASON PATTERN 3



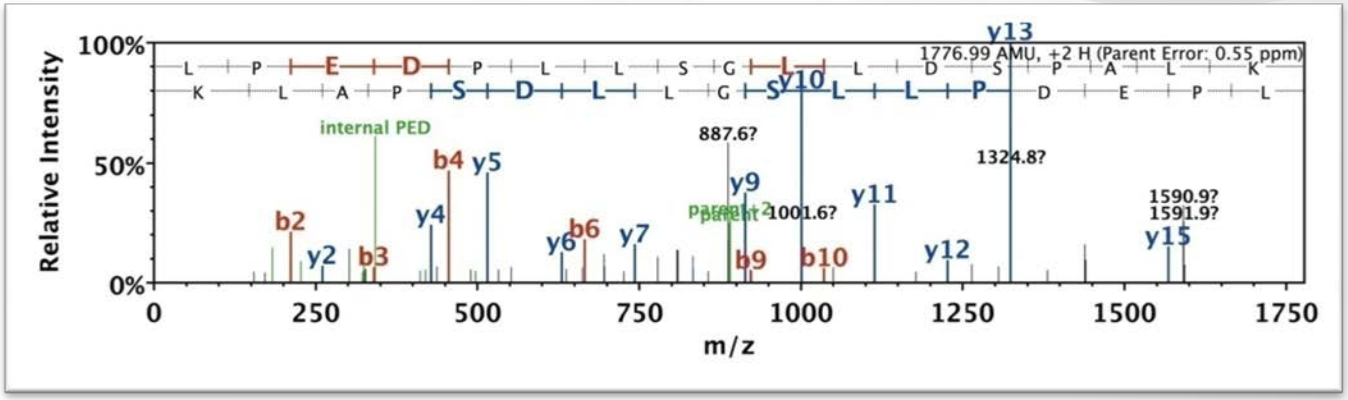


CGH DATAGENOMIC ANALYSIS

RESOLUTION = 10^{-7} to 10^{-9}

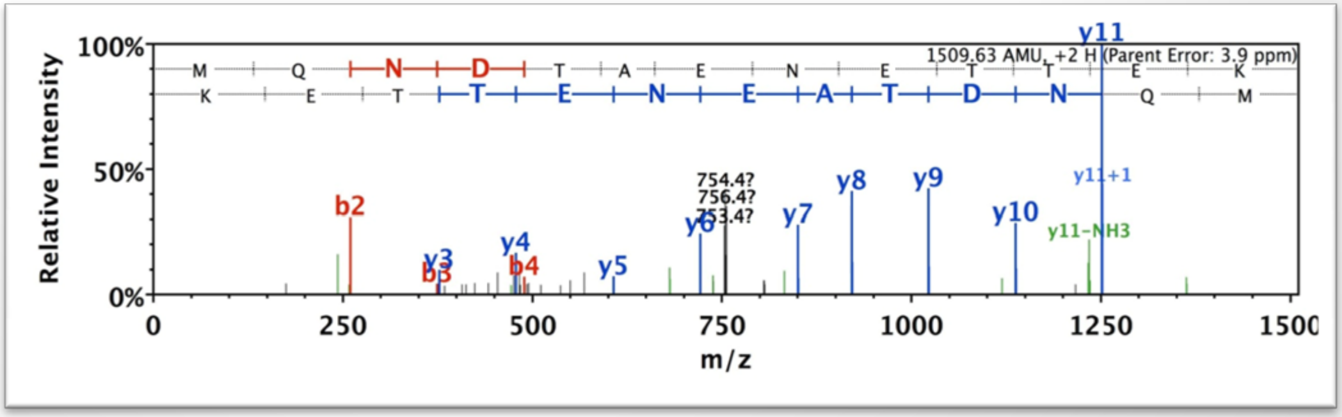
MS2 Spectra from Paraffin Tissue of CAP

Fatty Acid Synthase in CAP

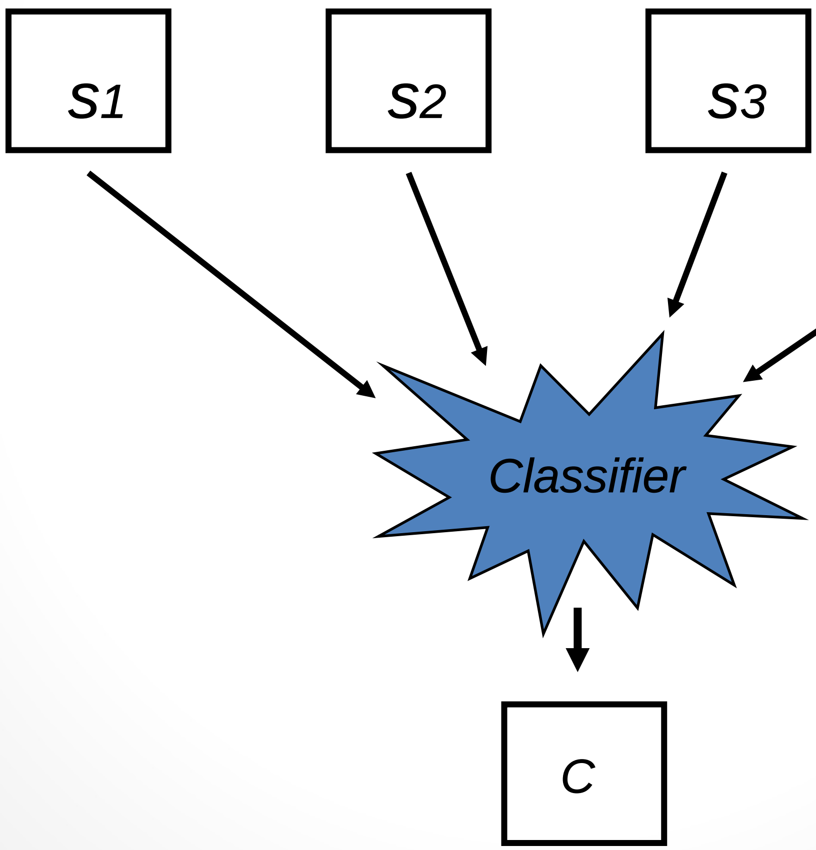


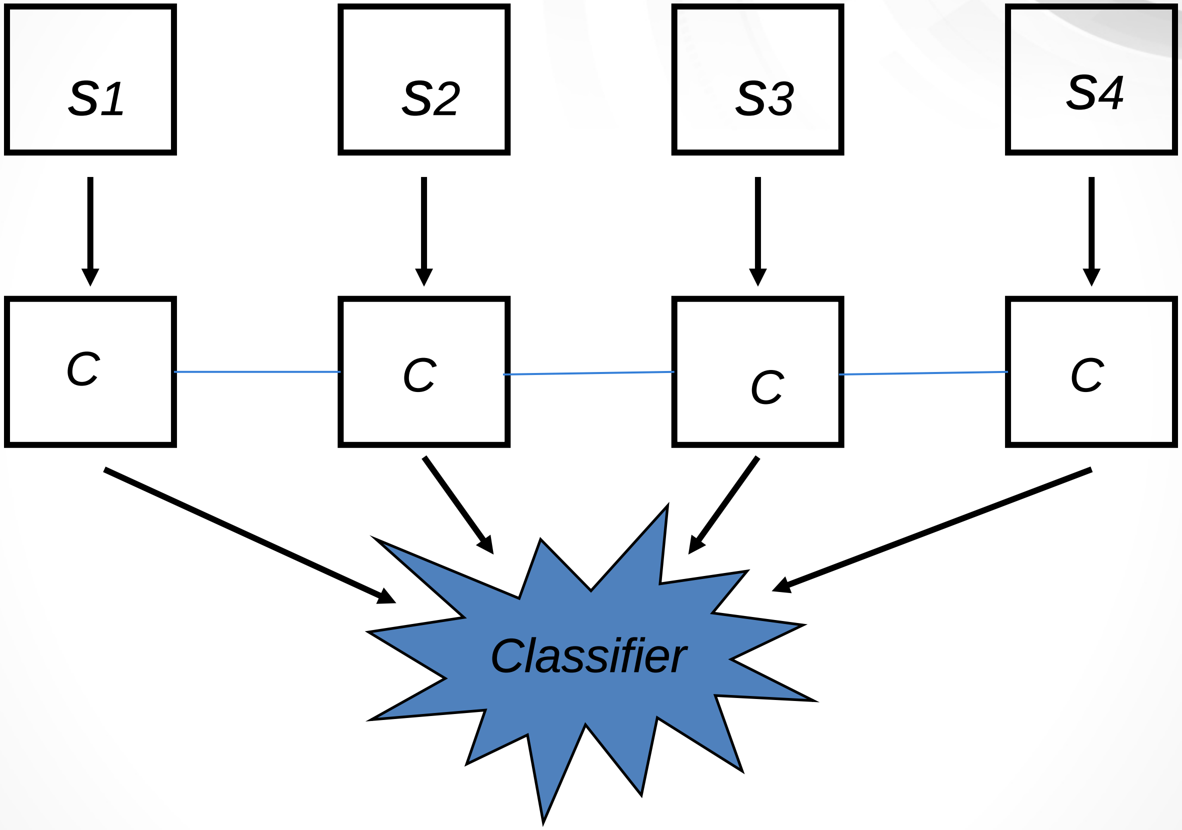
Mass Spec Protein Analysis *Resolution* = 10^{-10}

Caldesmon-1 in Benign Stromal Hyperplasia

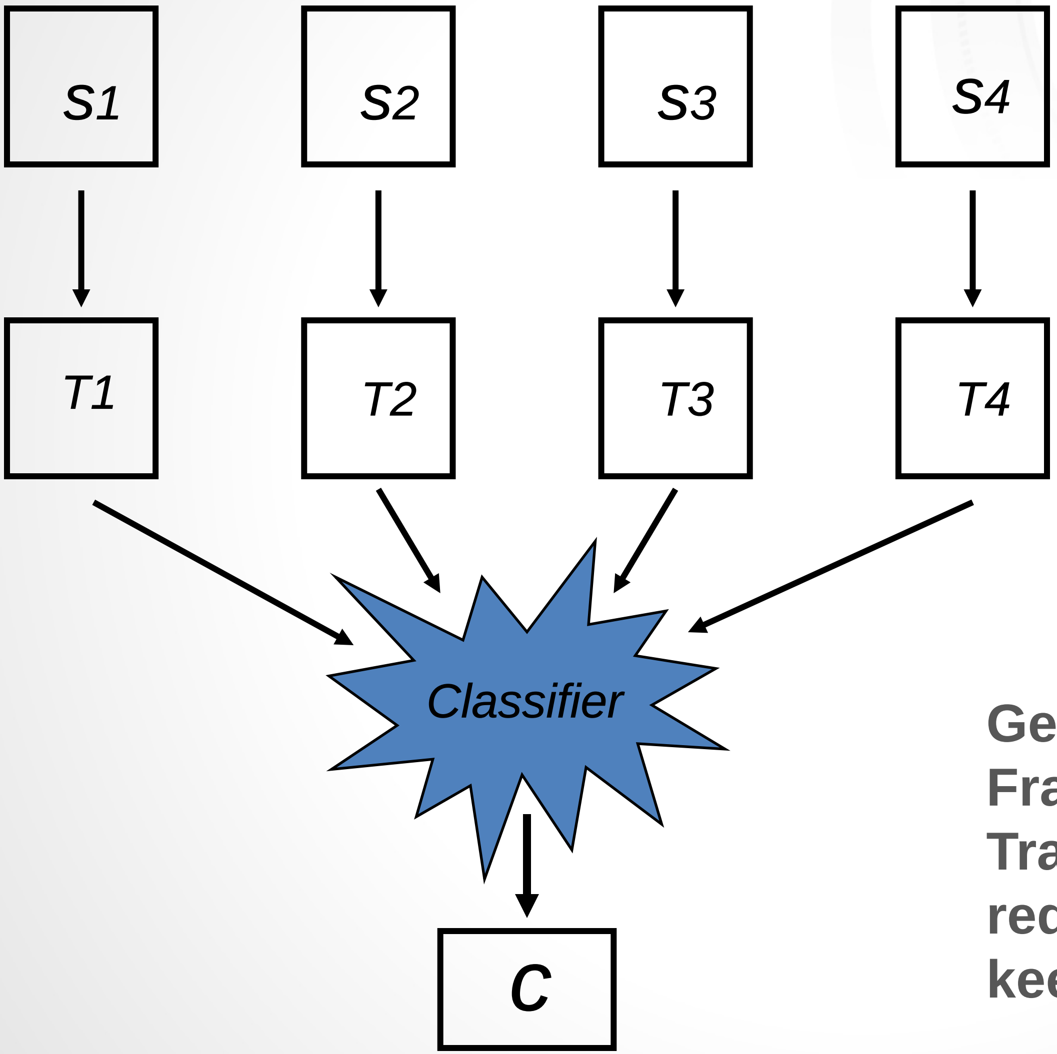


Combination of Interpretation



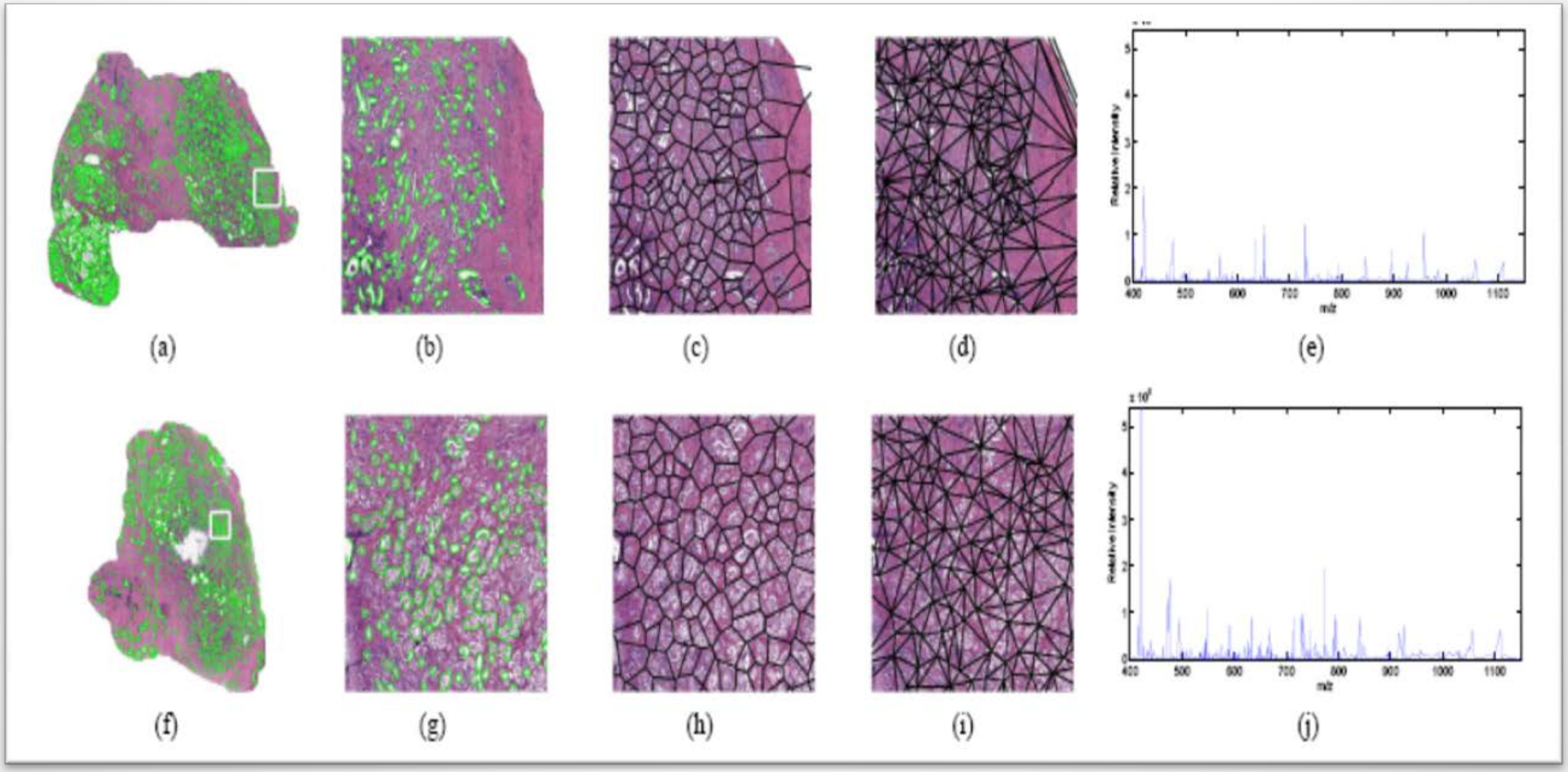


Combination of Data

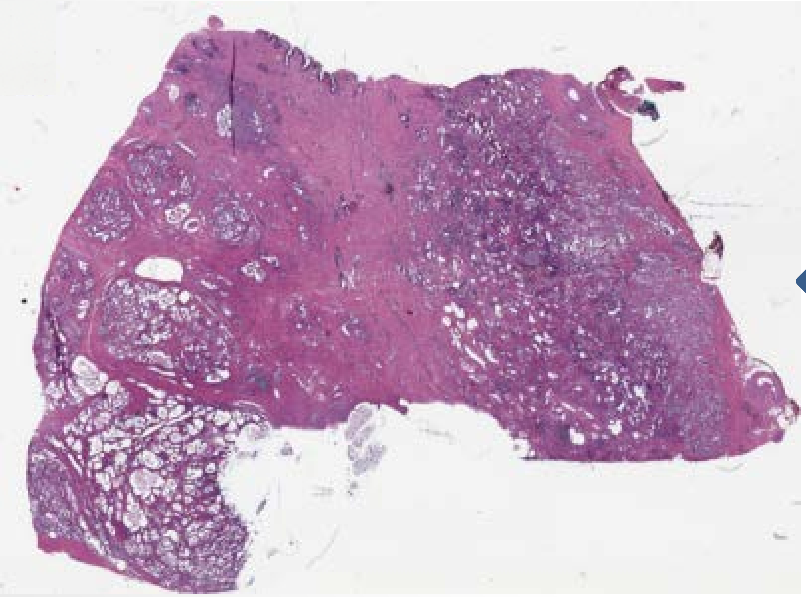


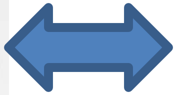
Generalized Fusion Framework = Transforms of data which reduce dimensions but keep links before classifier

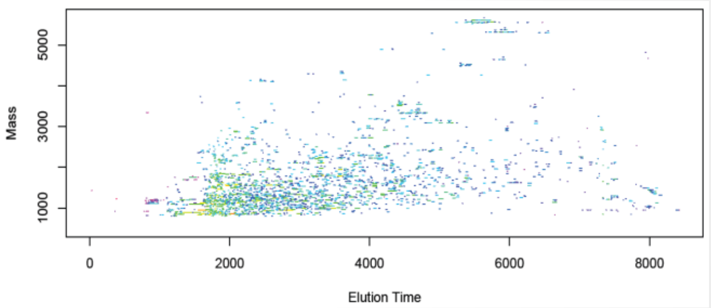
Multimodal Data Converge:
A General Fusion Framework Pilot Data



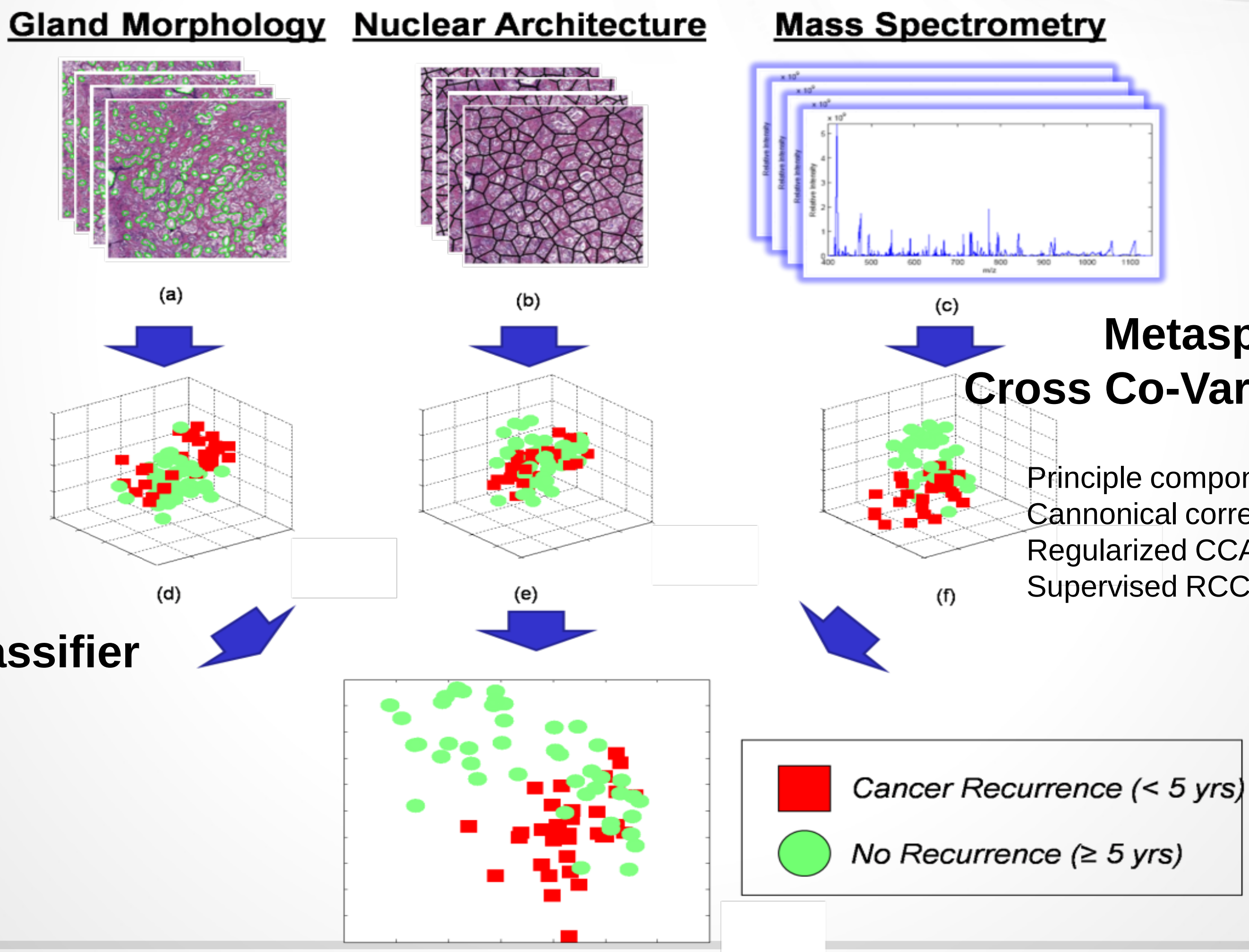
- **Aim**: Fuse image and proteomic features to identify patients with high risk of recurrence







Generalized Fusion Framework

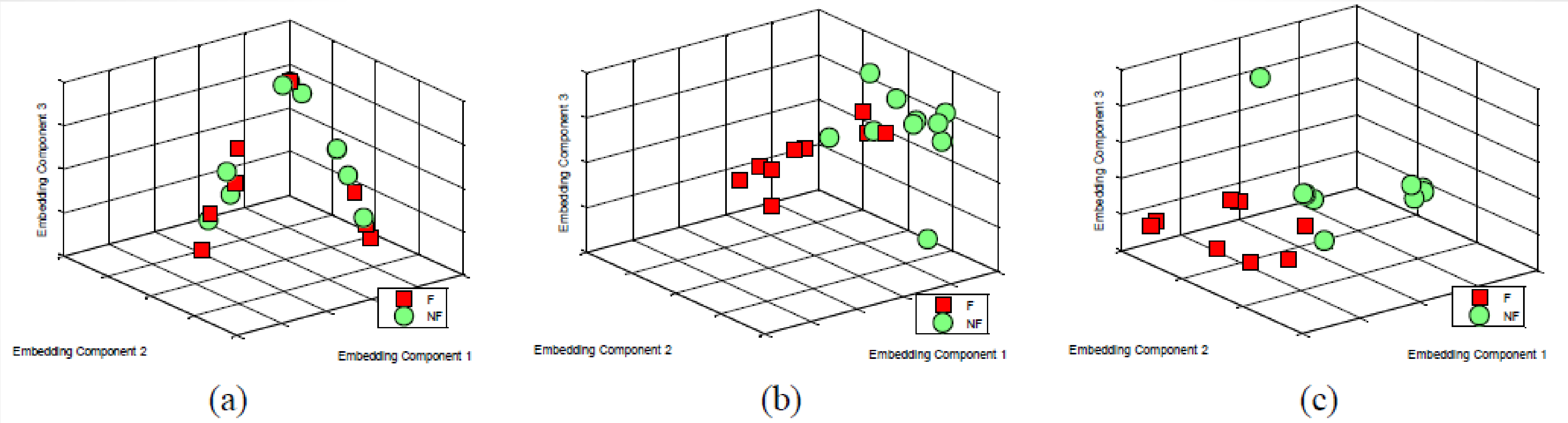


Metaspace via Cross Co-Variance Matrices

Principle component analysis (PCA)
Cannonical correlation analysis (CCA)
Regularized CCA (RCCA)
Supervised RCCA

Classifier

Visualizing Graph, Protein, Fused Data



Graph-based Features

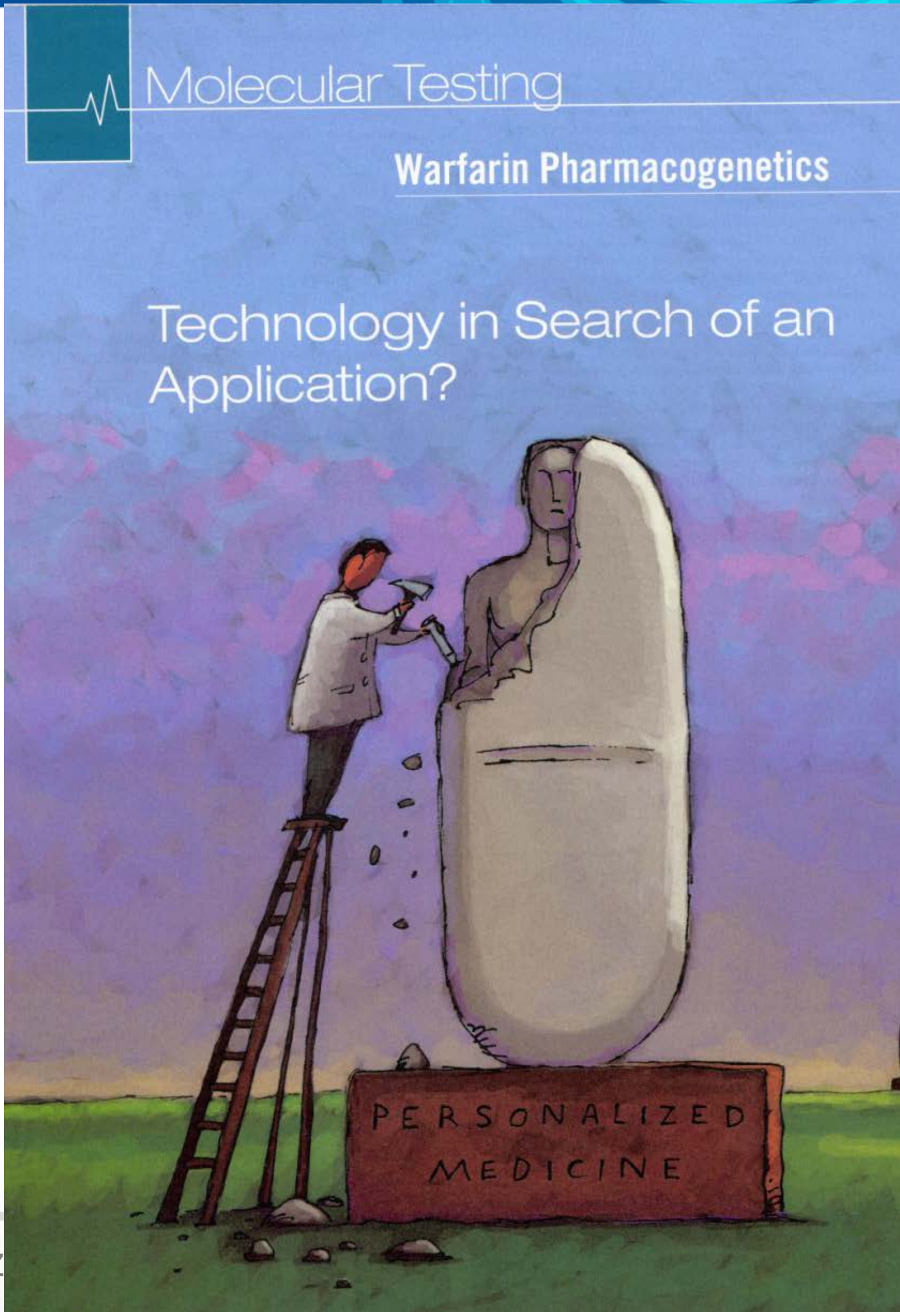
Proteomic Features

Fused Feature Space

BMC Bioinformatics 2011, 12:483

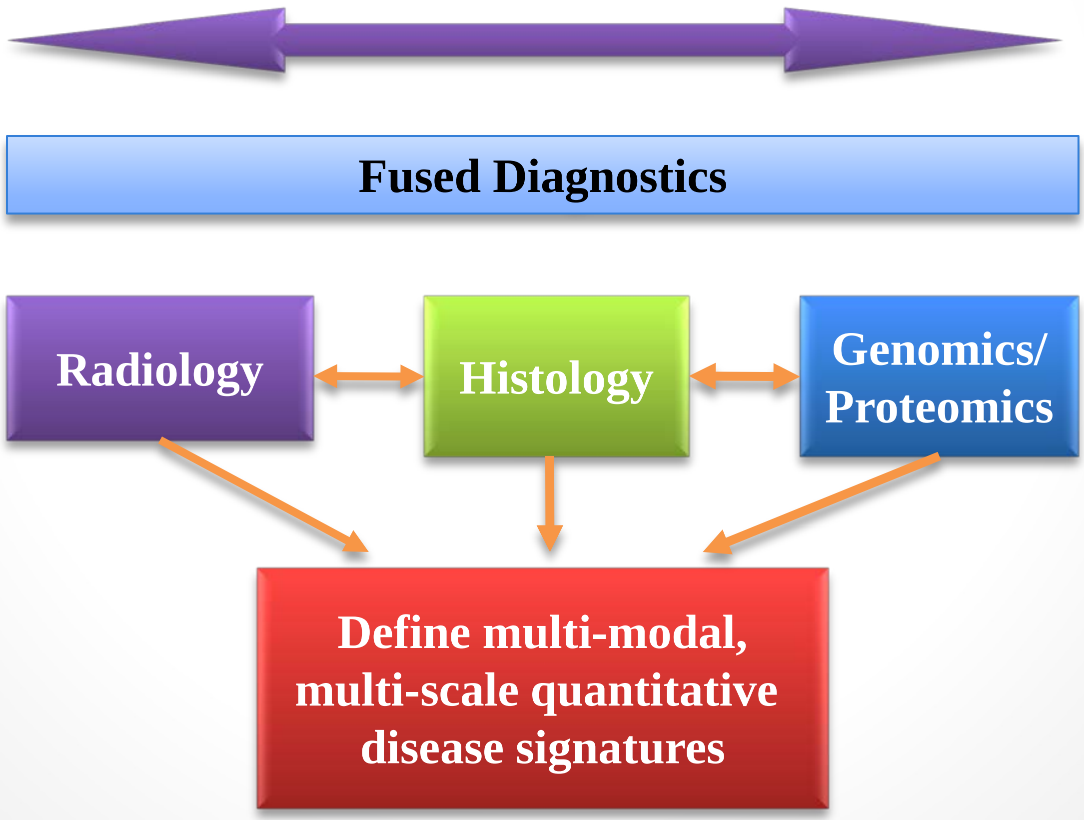
Chiseling out Pathology's Place in 21st Century Diagnostics

(it's about the analytics)



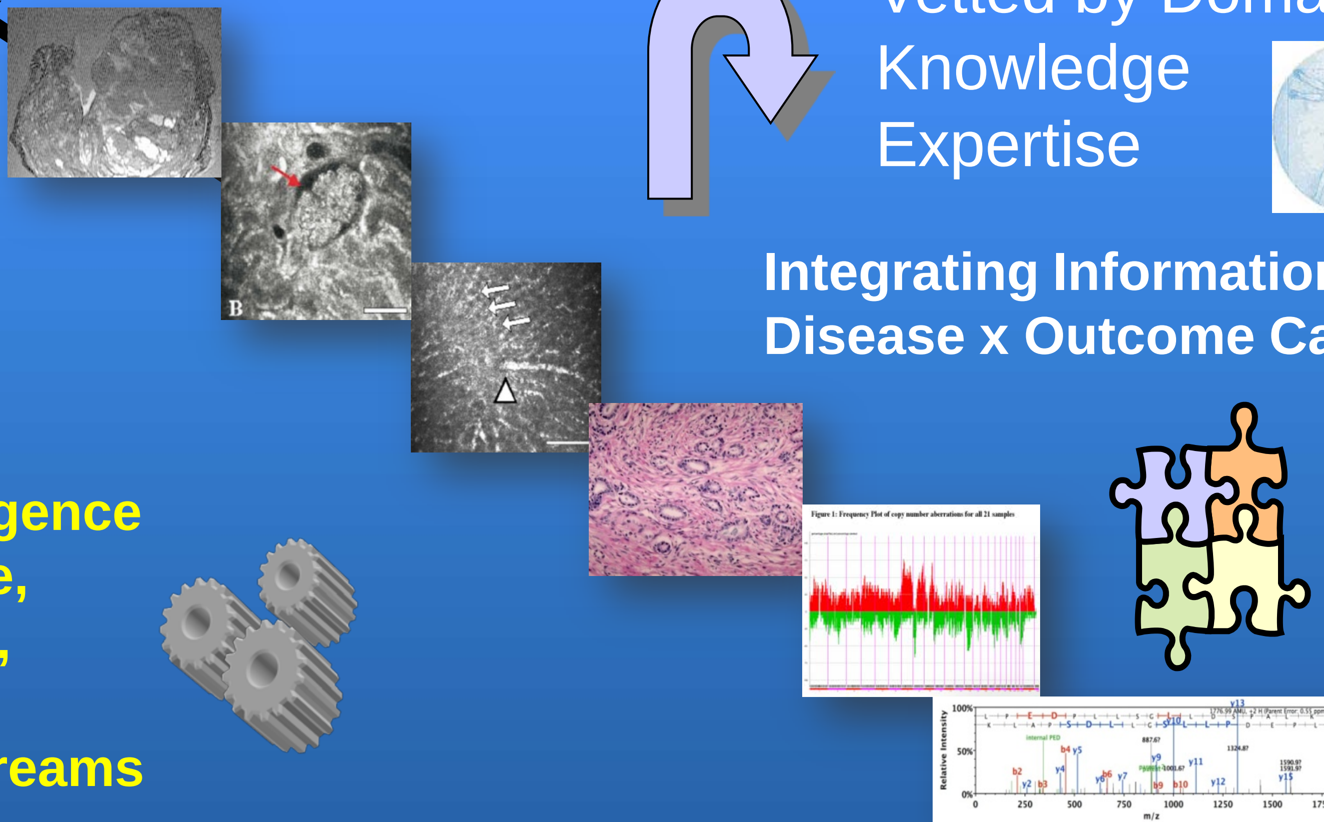
John E. Tomasz

Core Philosophy
Multi-modal data Integration
Personalized Medicine Form 3 =



Diagnostic Imaging in 2020

10^{1}



Vetted by Domain Knowledge Expertise

Integrating Information in Disease x Outcome Categories

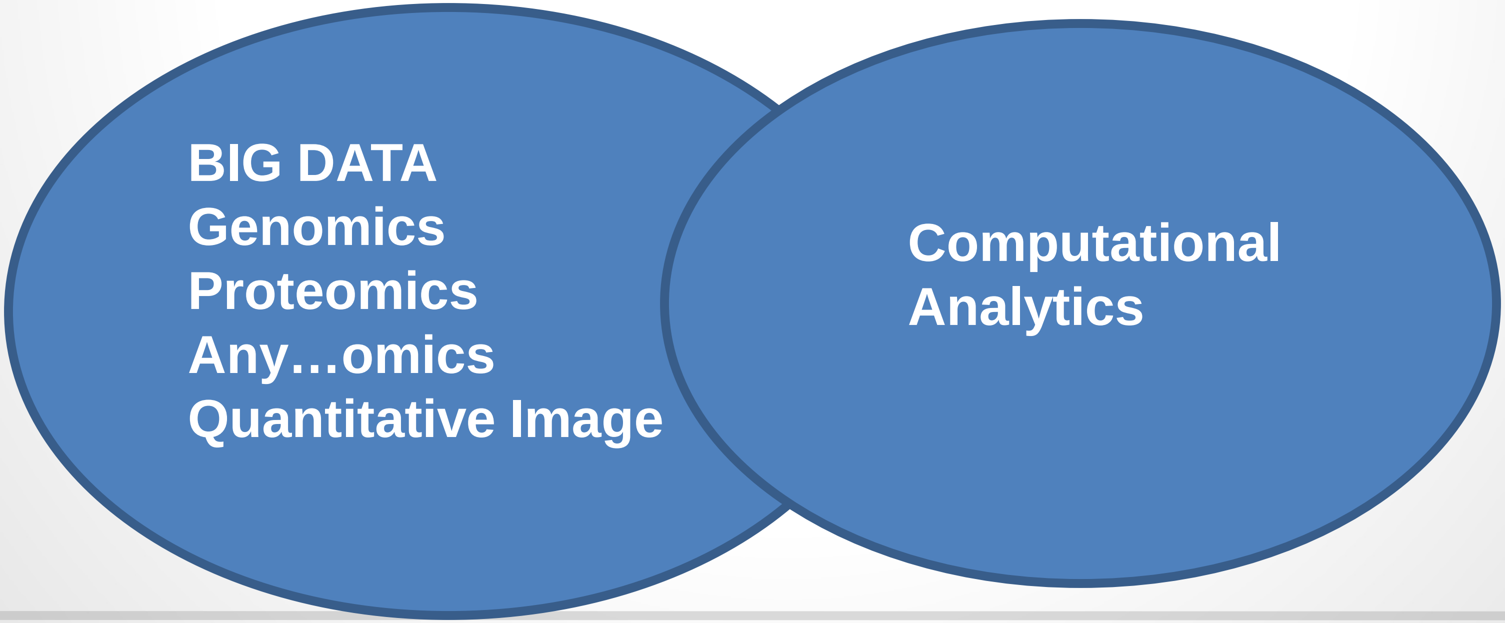
Convergence of Large, Diverse, Scaled Data Streams

10^{-10}

Facilitated by Machine Learning

Fused Diagnostics
Integrated Diagnostics
Advanced Analytics
Computational Diagnostics

Personalized Predictive Modeling
(Precision Diagnostics)



Evolutionary Implications of "Precision Diagnostics" for the Practice of Pathology

- Personalized Medicine/Molecular Diagnostics
- Digital Pathology/Telepathology
- Consultative Role
- Reward/Pay for Reduction in Unnecessary Testing
- Diagnostic Wing of the Medical Home
- Direct Patient Care

"Change is the law of life. And those who look only to the past or present are certain to miss the future."

John F. Kennedy

Thank You !