

Integrative Genomics Approach for Precision Medicine in Cancer

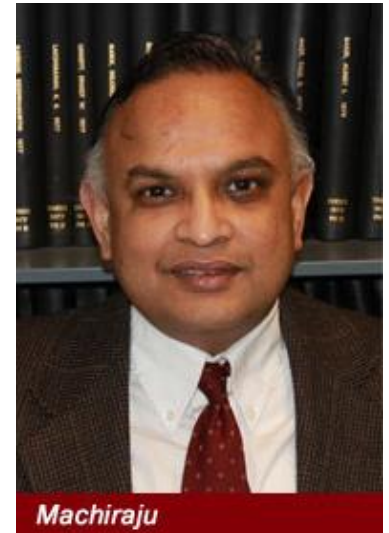
Kun Huang

**Department of Medicine
Director for Data Science and Informatics
Precision Health Initiative
Indiana University School of Medicine
Regenstrief Institute**

October 17, 2017

Outline

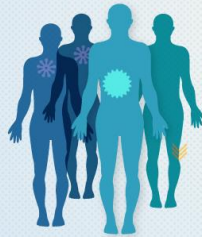
- **Computational Pathology and Precision Medicine**
- **What Information Does Imaging Data Provide - Quantitative Phenotyping Tool**
- **Complementary with Omics Data - Integrative Genomics**
- **Summary and Perspective**



Precision Medicine and Data

NATIONAL CANCER INSTITUTE PRECISION MEDICINE IN CANCER TREATMENT

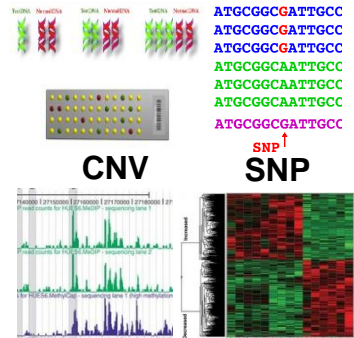
Discovering unique therapies that treat an individual's cancer based on the specific genetic abnormalities of that person's tumor.



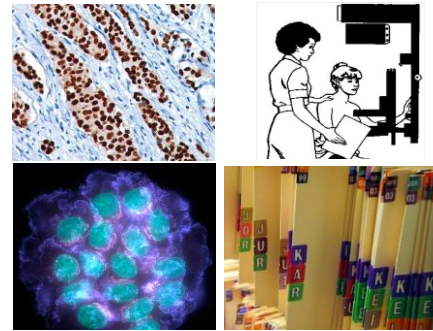
www.cancer.gov



Cancer Patients



Epigenome Gene Expression



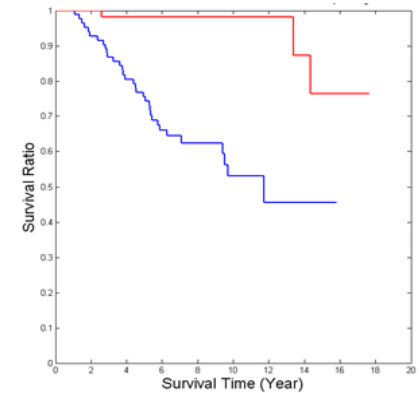
Clinical Phenotypes

Correlation / AI

Biomarkers for Clinical Prediction

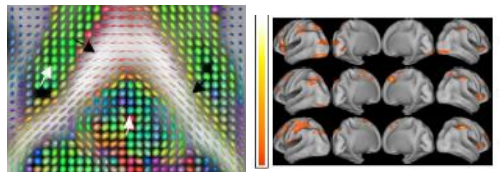
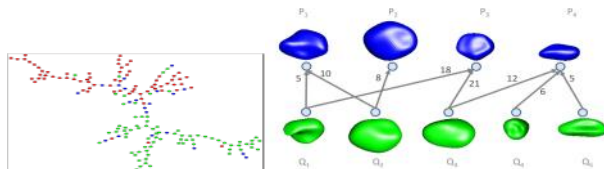
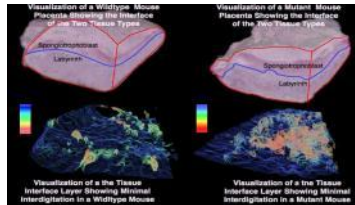
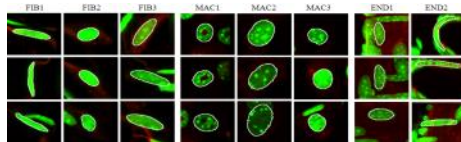
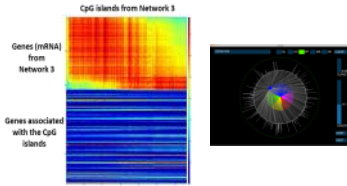


Patient Stratification



Big Data - Variety

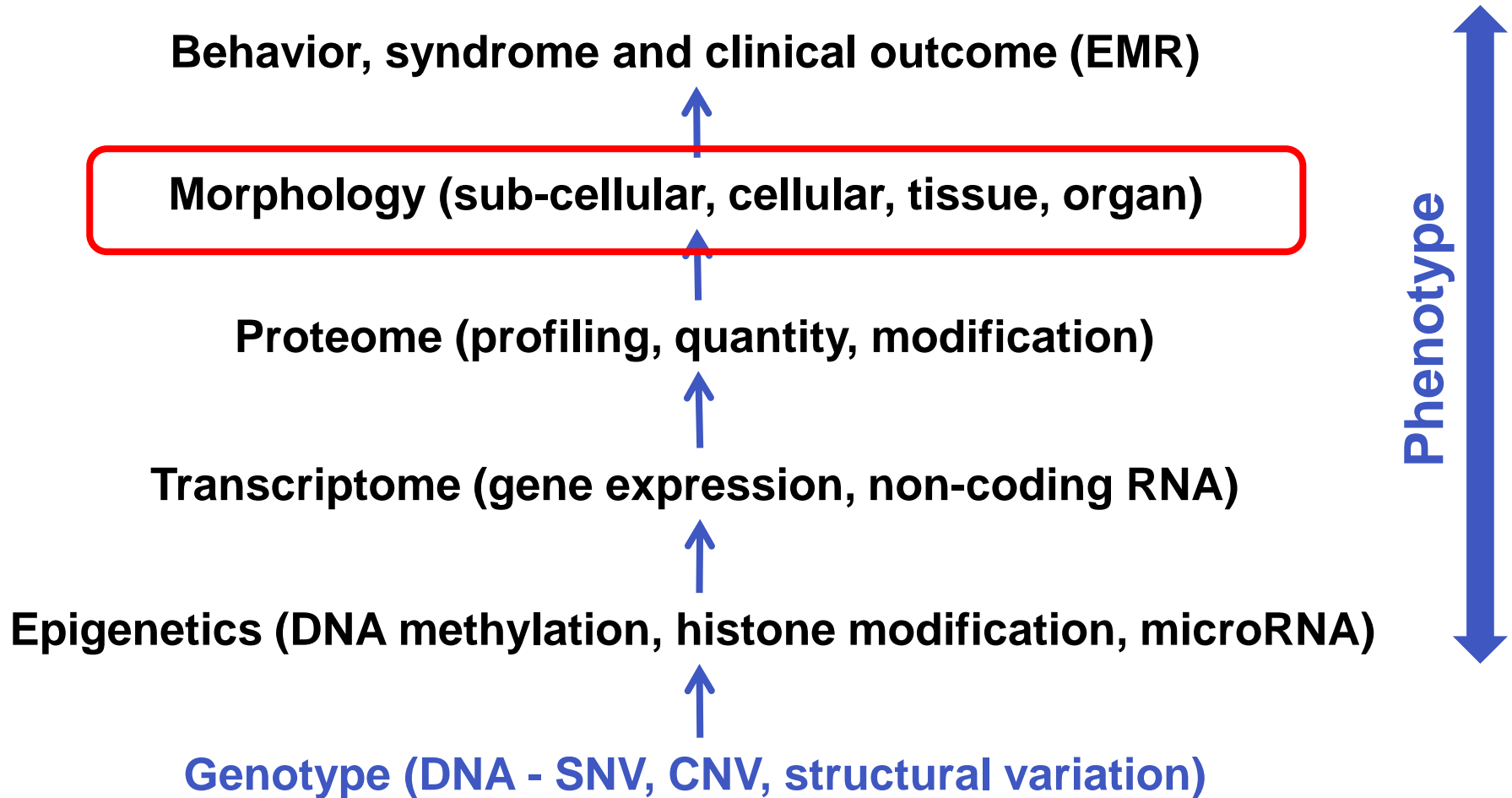
Multiple phenotypical scales



Level	Extensive	Intensive
DNA, RNA	Solved and costs declining rapidly. Limited by bioinformatics capabilities	Possible but costly; detailed sampling is challenging
Chromatin	Made possible by chromatin immunoprecipitation but with low precision; costly	Possible but costly; detailed sampling is challenging owing to instability
Proteins, metabolites	Mostly solved and capabilities are still improving for rare constituents	Possible but costly; detailed sampling is challenging
Cells	Huge diversity of assays is possible but most are low throughput. Image-based techniques for high throughput are promising	Many temporally continuous assays. Image-based approaches allow some combination of extensive and intensive measurement
Development	Low-throughput measures are abundant. High-throughput image-based methods are possible for small living individuals (embryos and <i>Caenorhabditis elegans</i>) and sectioned tissues	Temporal depth by repeated sampling or image-based approaches
Physiology	Huge diversity of assays is possible, but most are low throughput, unless based on proteomic or metabolomic data	Many temporally continuous assays. Spatial sampling is often possible
Morphology	Solved in principle with the use of imaging, but assays often require extensive sample preparation. Post-processing to extract and measure features requires specialized informatics capability	Intensive sampling of morphological form is possible with specialized processing. Temporal depth is limited by destructive sampling
Behaviour	Possible with continuous observation of video of confined or local populations, or for humans, self-reporting. Data extraction from video using specialized software or human labour	Extensive sampling gives intensive coverage. Specific aspects of behaviour can be intensively measured with data loggers

Data Integration

- Integrative genomics / trans-omics approach

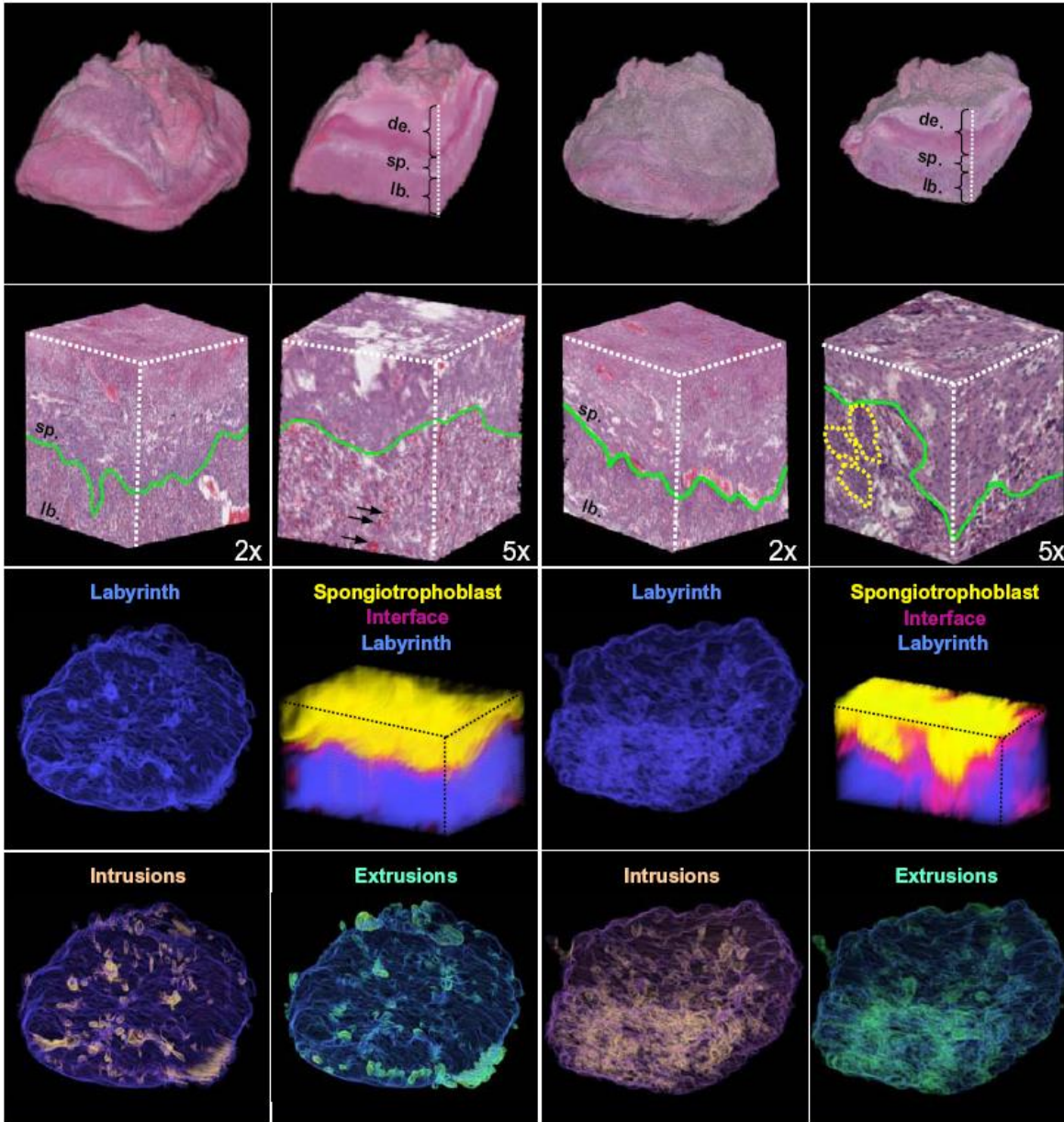


Outline

- **Computational Pathology and Precision Medicine**
- **What Information Does Imaging Data Provide - Quantitative Phenotyping Tool**
- **Complementary with Omics Data - Integrative Genomics**
- **Summary and Perspective**

$Rb^{+/+}$

$Rb^{-/-}$

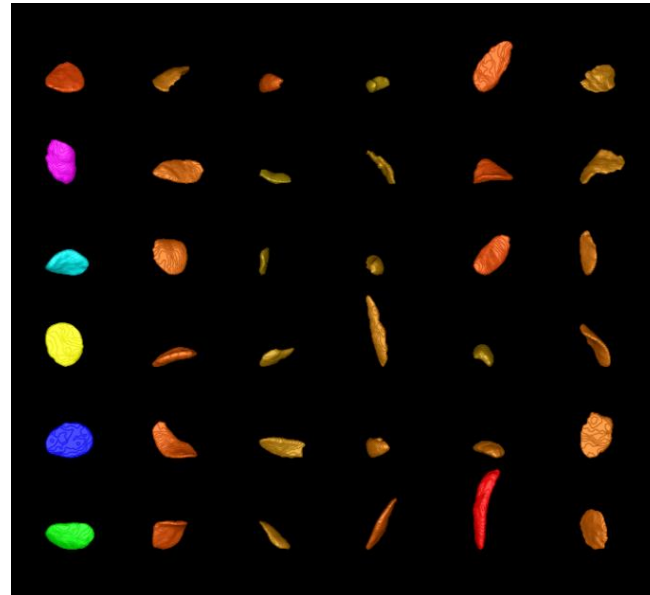
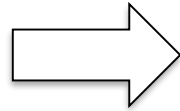
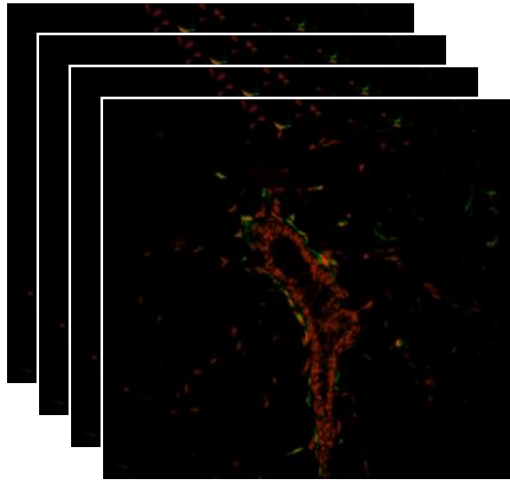
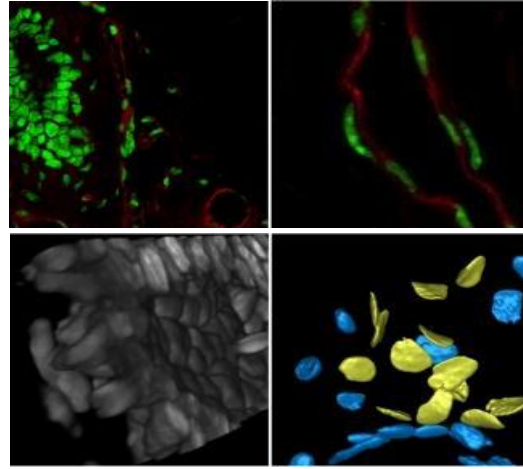
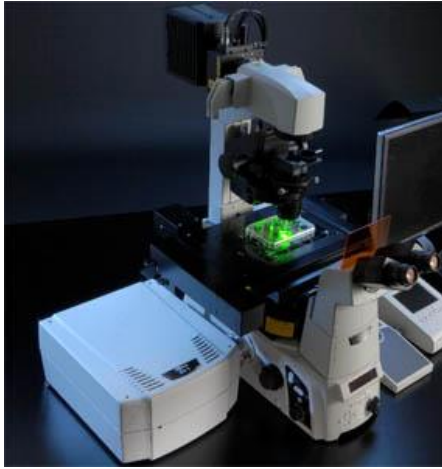


- Image registration
- 3D reconstruction
- Segmentation
- Visualization
- Quantification
- Computing



Wenzel et al, Gene & Development, 2007

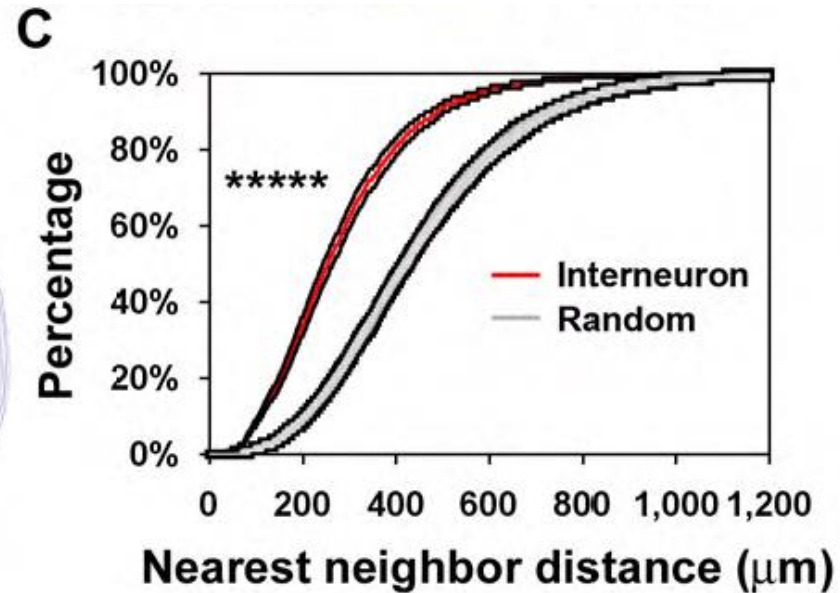
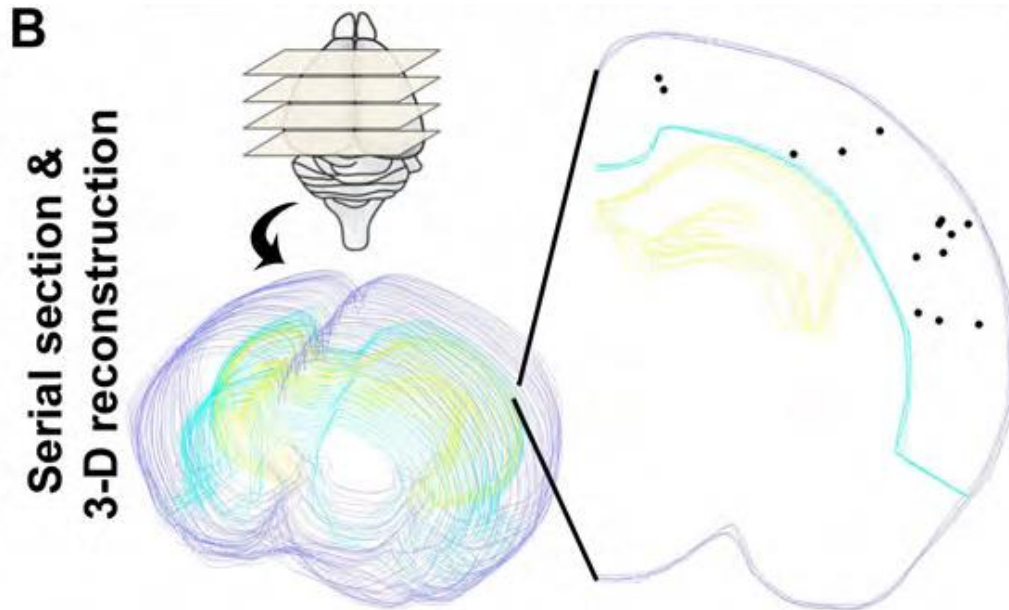
3D shape characterization using SPHARMS



Shantanu Singh
Broad Institute

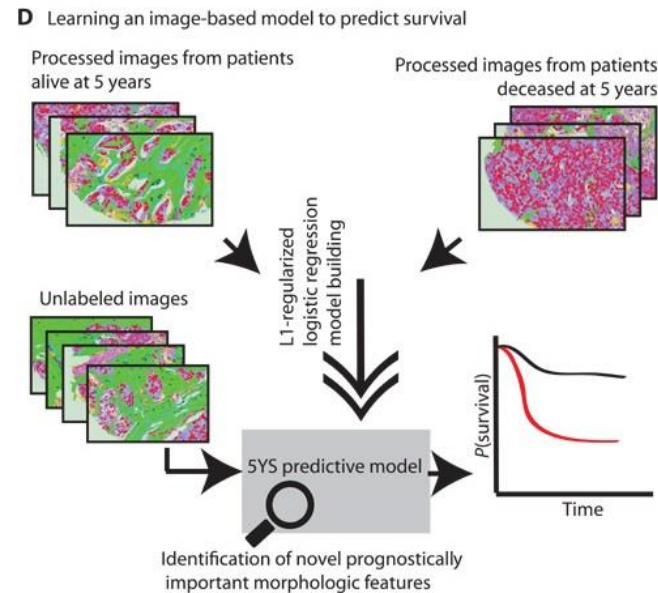
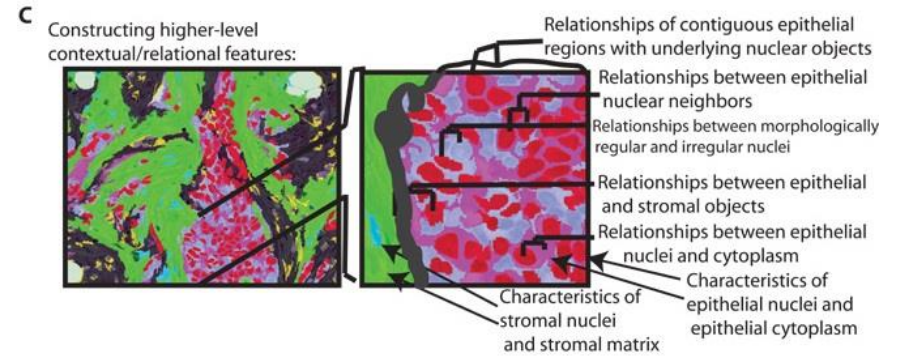
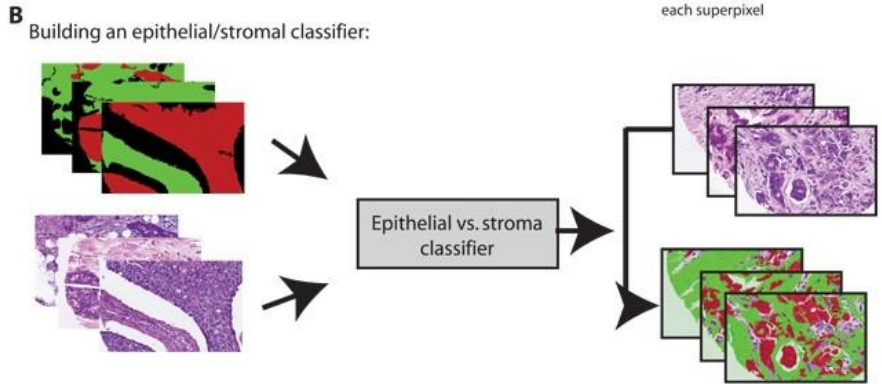
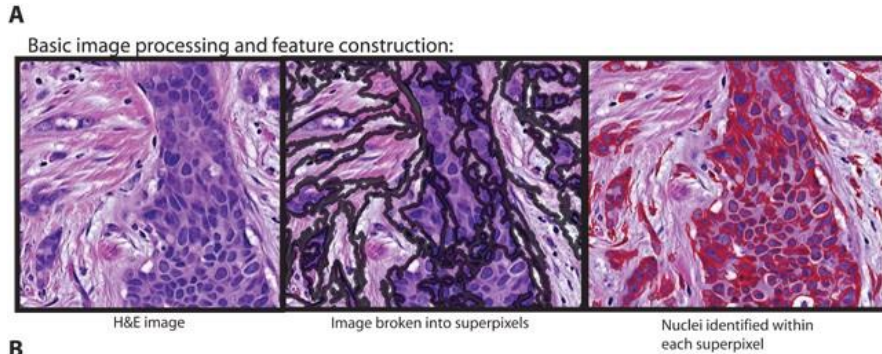
Neuroinformatics

- **Spatial pattern analysis**

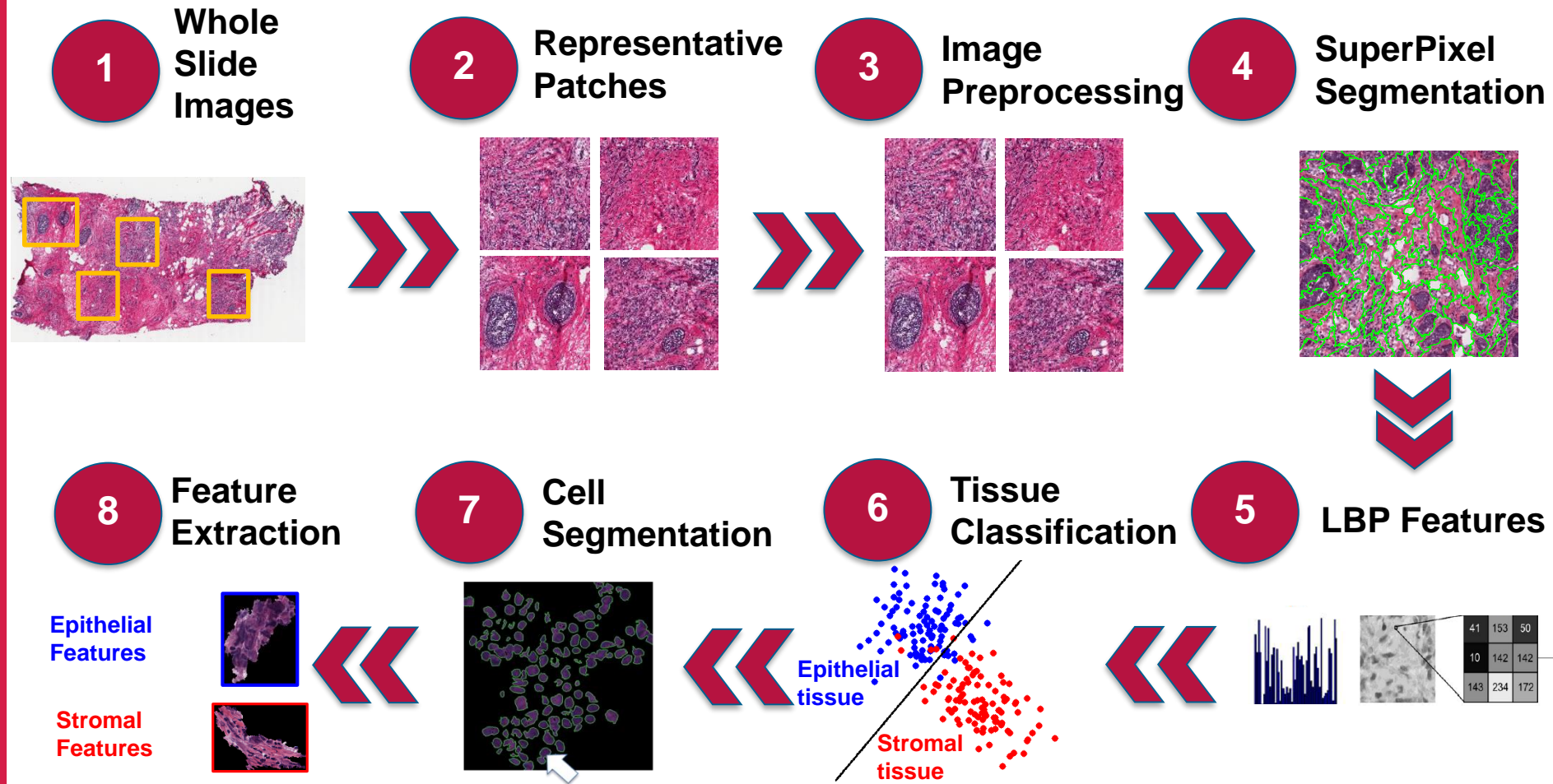


- **Brown KN, et al. Science, 2011.**
- **Yu Y-C, et al. Nature, 2012.**
- **Xu H-T, et al. Cell, 2014.**
- **Gao P, et al. Cell, 2014.**
- **Sultan S, et al. Neuron, 2016.**
- **Shi W, et al. Nature Neuroscience, 2017.**
- **Zhang X, et al. Nature Communications, 2017.**

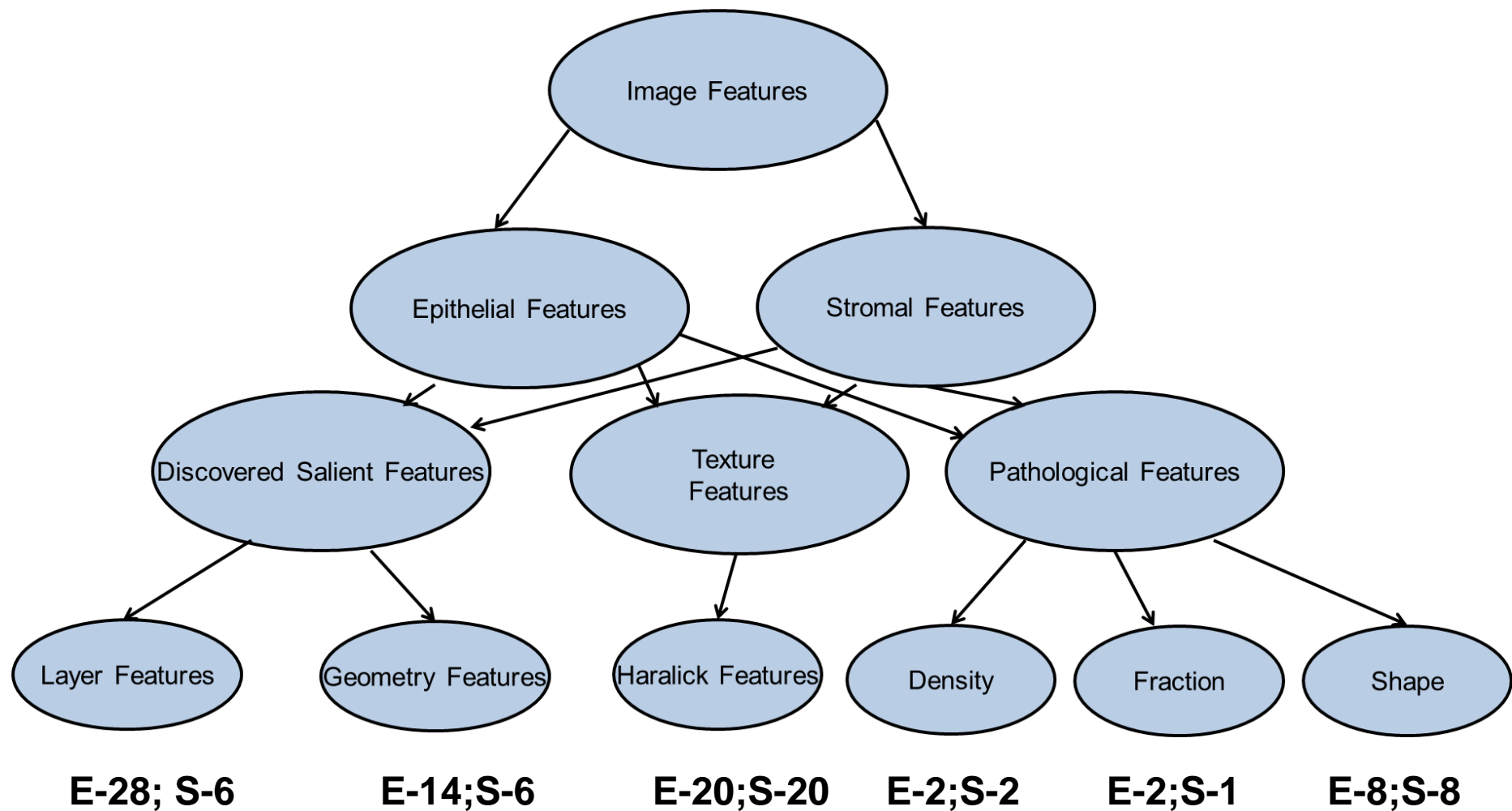
Histopathology Features for Cancer Prognosis



Pipeline Overview

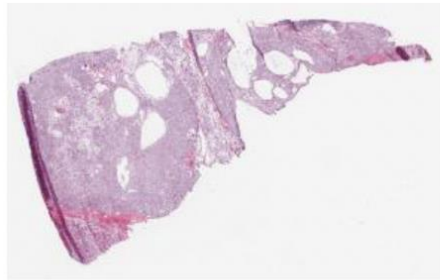


Feature Extraction

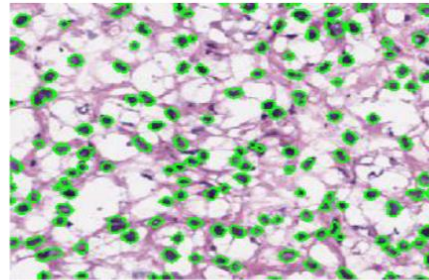


E:epithelial features;S:stromal features

Feature Distribution



Whole-slide image



Nucleus segmentation



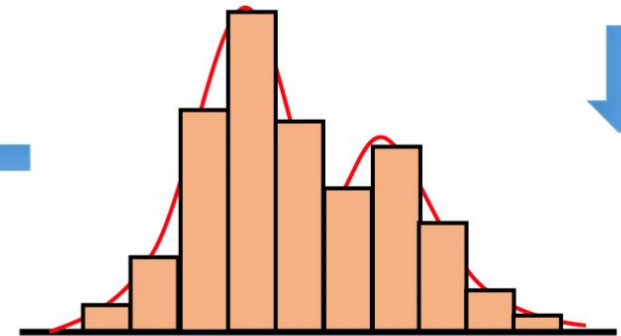
Area
Major axis length
Minor axis length
Ratio ...

Cell-level features



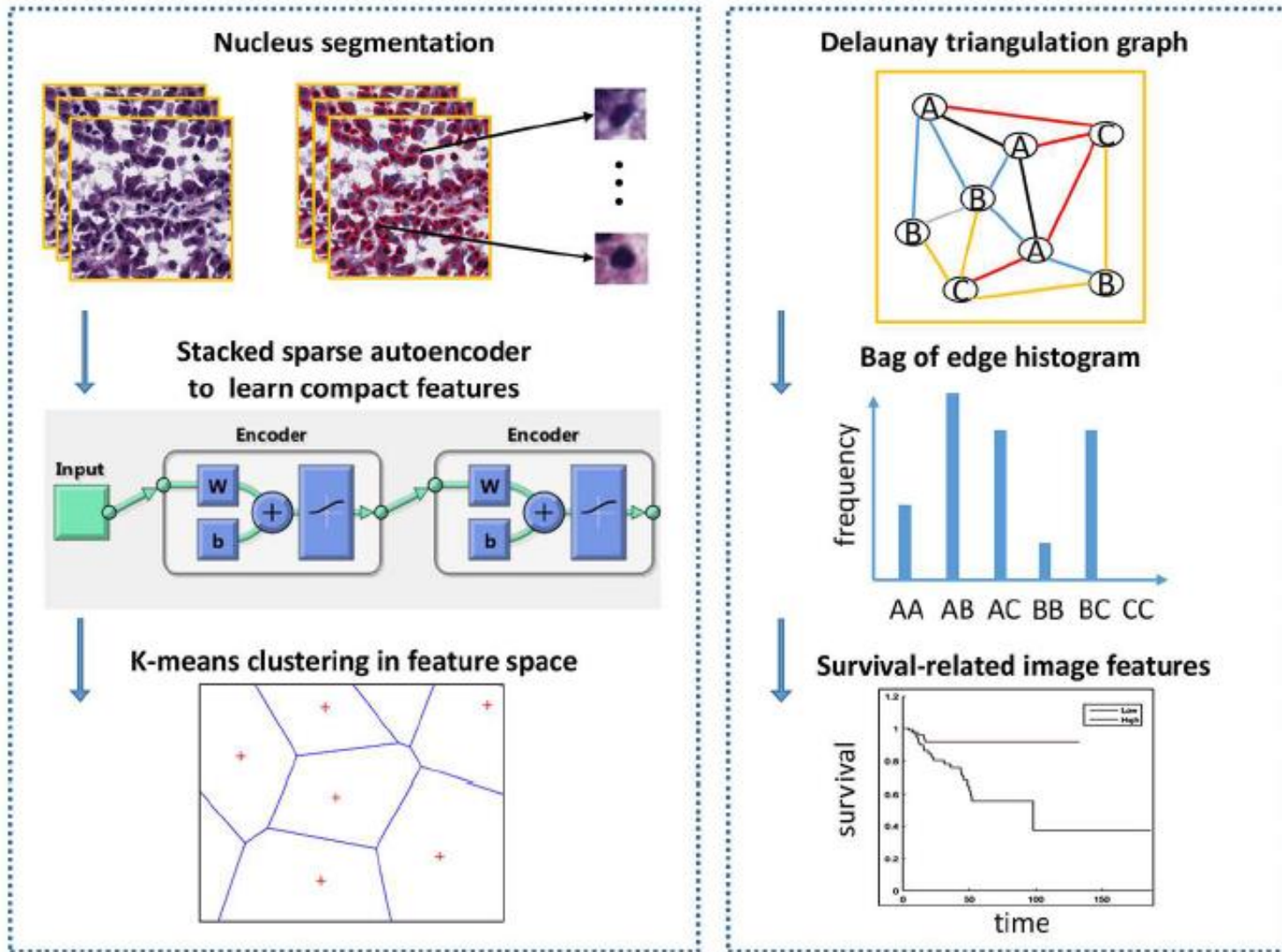
❖ Histogram
10-bin histogram
❖ Distribution statistics
Mean, standard deviation,
skewness, kurtosis, entropy

Patient-level features



Area

Topological Features – Kidney Cancer



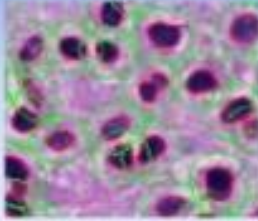
Cheng et al, 2017

Outline

- **Computational Pathology and Precision Medicine**
- **What Information Does Imaging Data Provide - Quantitative Phenotyping Tool**
- **Complementary with Omics Data - Integrative Genomics**
- **Summary and Perspective**

Morphology engine

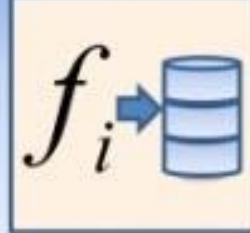
Segmentation



Feature extraction



PAIS database

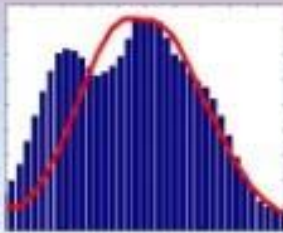


Patient modeling

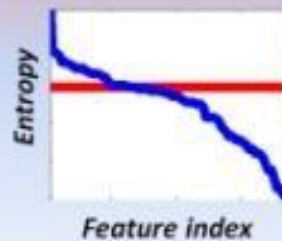


Clustering engine

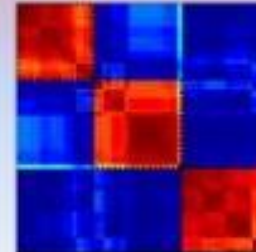
Normalization



Feature selection

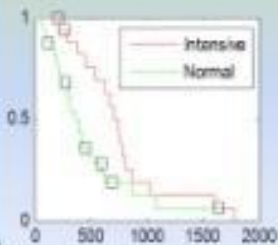


Consensus clustering



Correlative engine

Survival analysis



Molecular classes

Proneural
Classical
Mesenchymal
Proliferative
GCIMP+

Human pathology

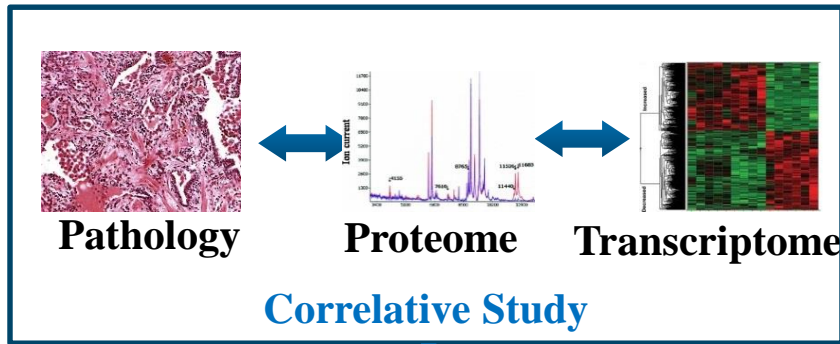


Genetic alterations

TP53 +/-
EGFR Amp.
CDKN2A Del.
⋮

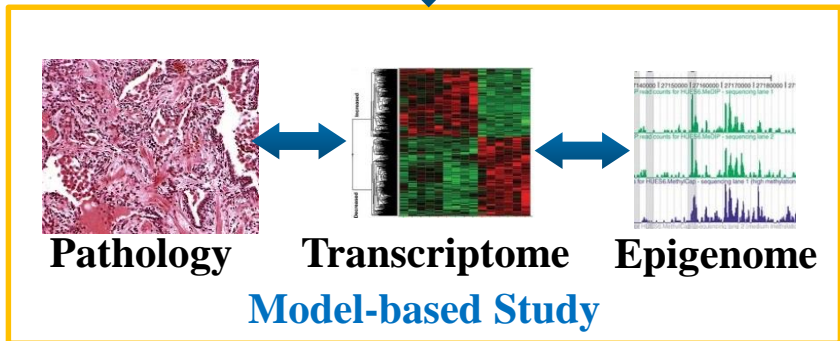
Cooper L et al, JAMIA, 19(2), 2012.

From Correlation to Integration



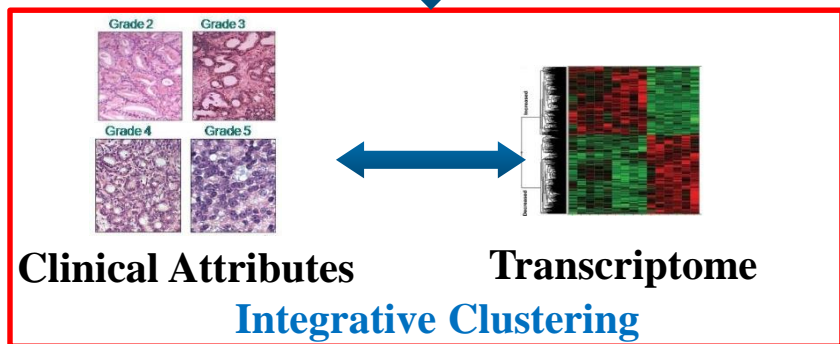
- Triple Negative Breast Cancer
- Breast Cancer Proteomics

Applications



- Lung Adenocarcinoma

Applications



Breast Cancer Patient Stratification

- A two-step algorithm
- Molecular regularized algorithm

Applications

Triple Negative Breast Cancer

TCGA

- 52 TNBC samples
- 154 images(40X)
- mRNA transcripts from RNA-seq

OSU TNBC TMA

- 173 TNBC samples
- 173 images(20X)
- 200+ genes

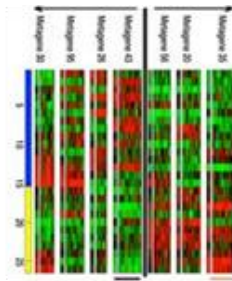
Public Gene Expression Datasets

- NKI
- Perou
- Wang

Overall Workflow

A TCGA

Gene expression profiles of these tissues



Metagenes

Correlation map



Candidate Correlated image features and genes pairs

Histopathological Images

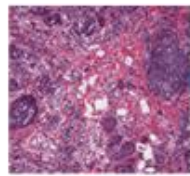
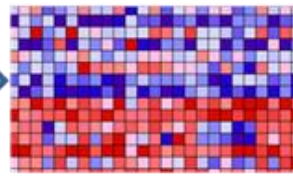
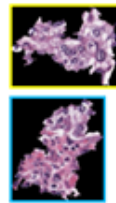
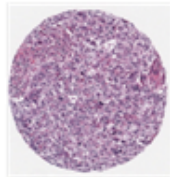


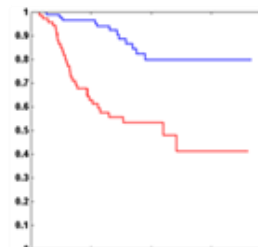
Image features



B OSU TNBC TMA



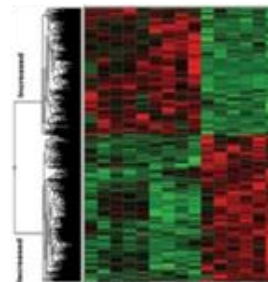
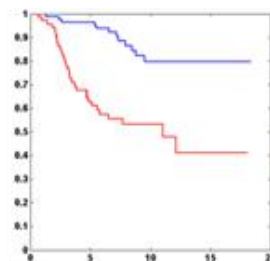
TMA image features



Survival-related image features

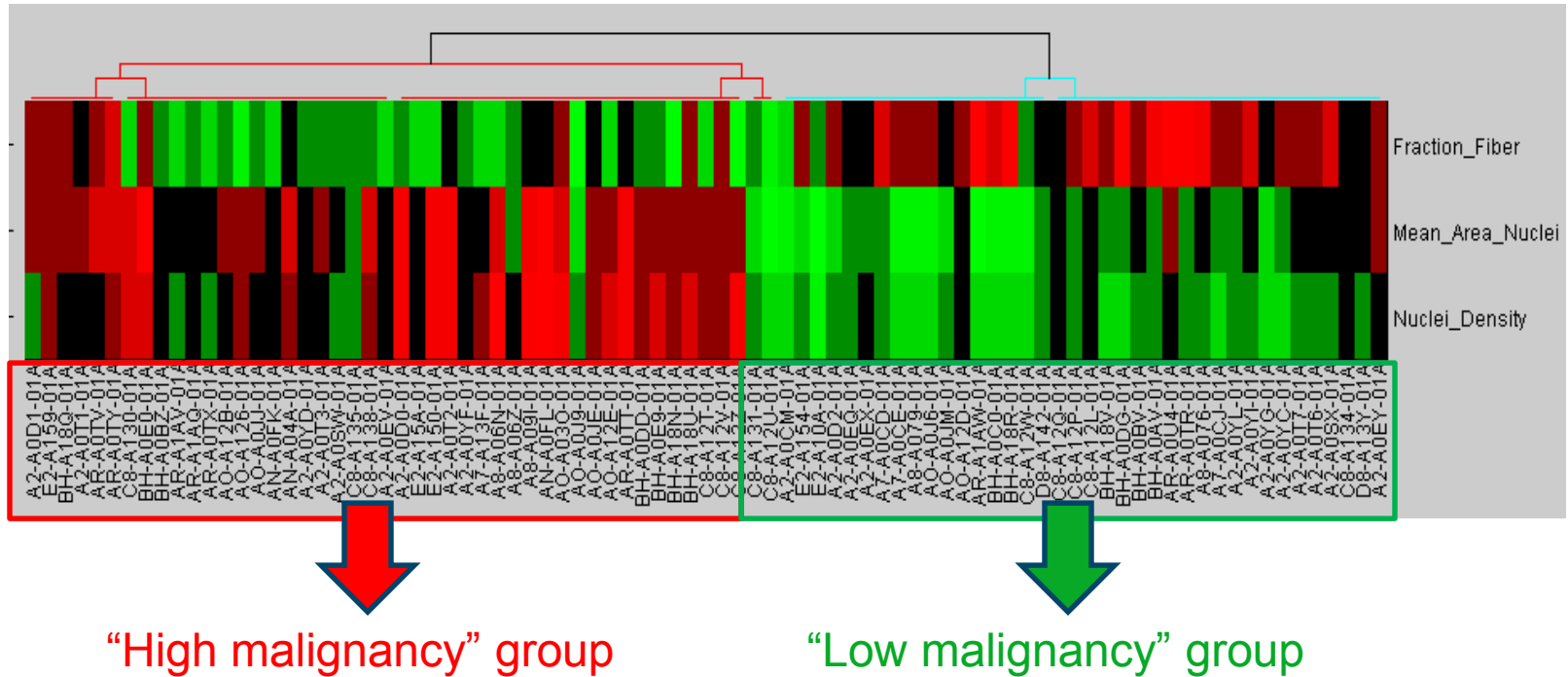
C

Multiple public breast cancer gene expression data with survival



Identify gene lists/clusters with strong correlation

Patient Stratification on Morphology



Hierarchical clustering of breast cancer patients based on imaging features.

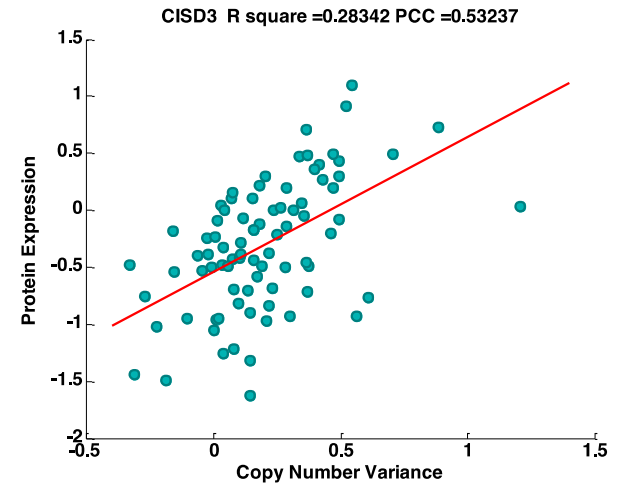
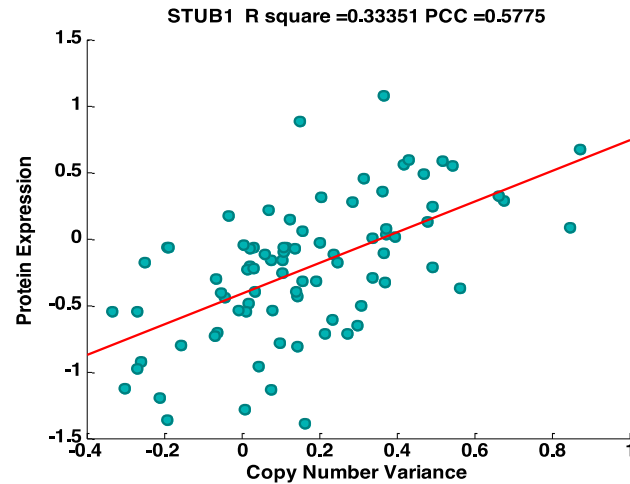
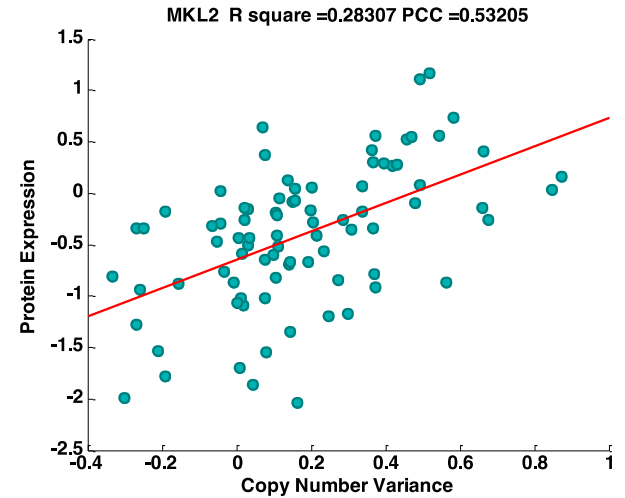
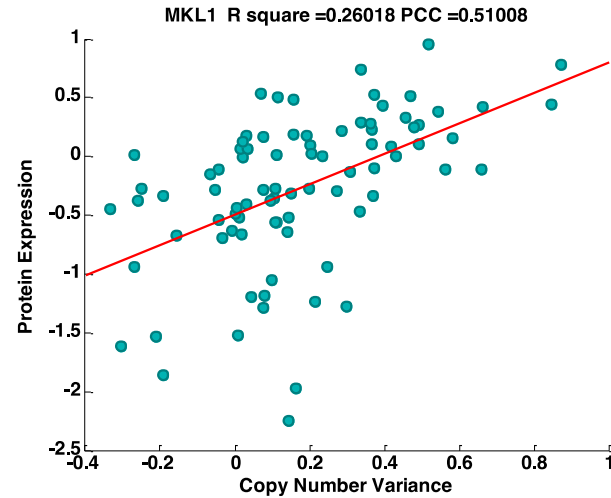
Associated Protein Co-Expression Network

- Identified 124 protein co-expression networks
- 4 protein networks differentially expressed

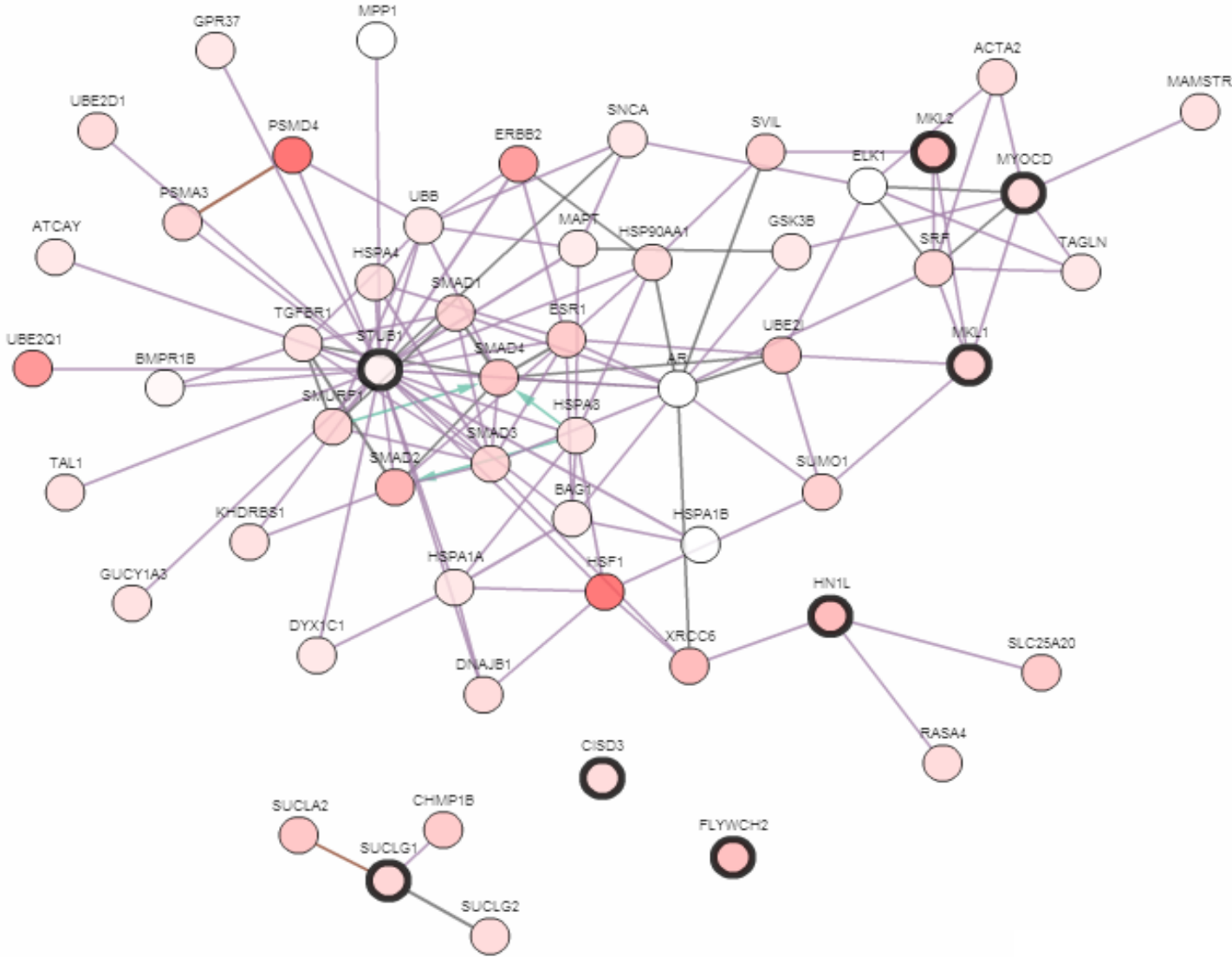
Protein Networks	Proteins in the Networks	Cytoband	pValue	Enriched Cytobands	Top GO Molecular Functions (pValue)	Note
1	MYOCD	17p11.2	3.880E-2	16p13 P – value = 2.170E-5	smooth muscle cell differentiation(8.184 E-7); muscle cell differentiation(1.889 E-4); regulation of transforming growth factor beta receptor signaling pathway.	Enriched in 16p13
	MKL1	22q13	5.986E-3			
	MKL2	16p13.12	2.306E-3			
	FLYWCH2	16p13.3	2.170E-5			
	CISD3	17q12	1.439E-1			
	HN1L	16p13.3	2.170E-5			
	STUB1	16p13.3	2.170E-5			
	SUCLG1	2p11.2	1.536E-2			

DNA Copy Number Variance

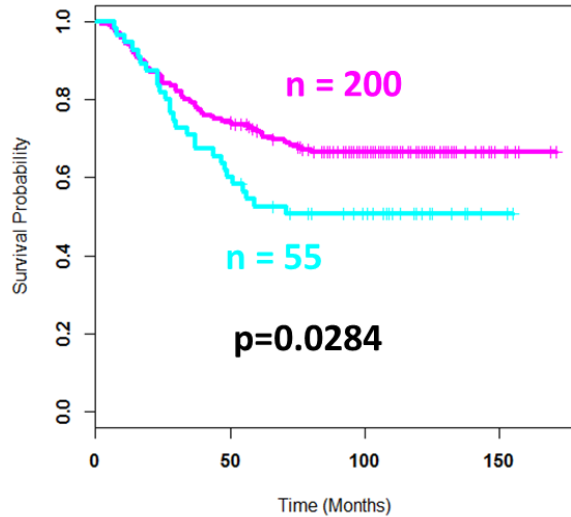
Protein Networks	Proteins in the Networks
1	MYOCD
	MKL1
	MKL2
	FLYWCH2
	CISD3
	HN1L
	STUB1
	SUCLG1



STUB1 gene as a hub gene

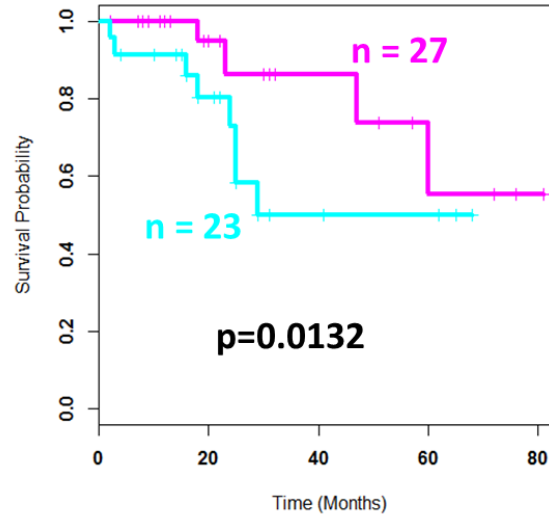


Prognosis Validation



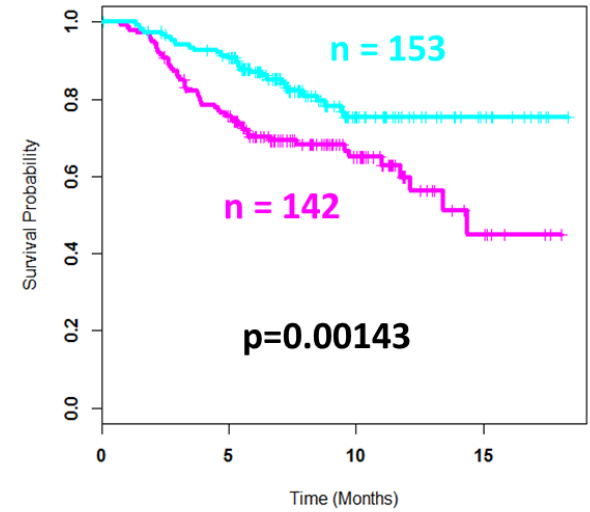
A

Wang



B

Perou



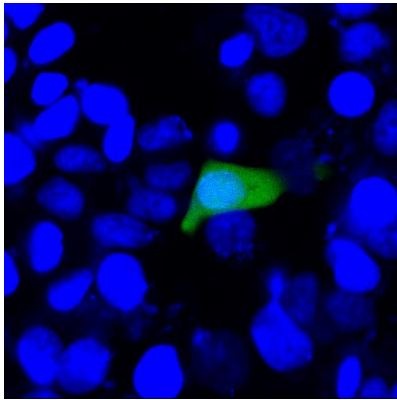
C

NKI

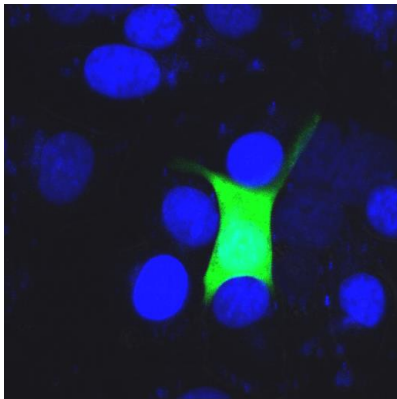
Kaplan–Meier survival curves of prognostic model on multiple public breast cancer datasets. From left to right: Wang, Perou and NKI data respectively.

STUB1 Knockout in MCF7 Breast Cancer Cell Line

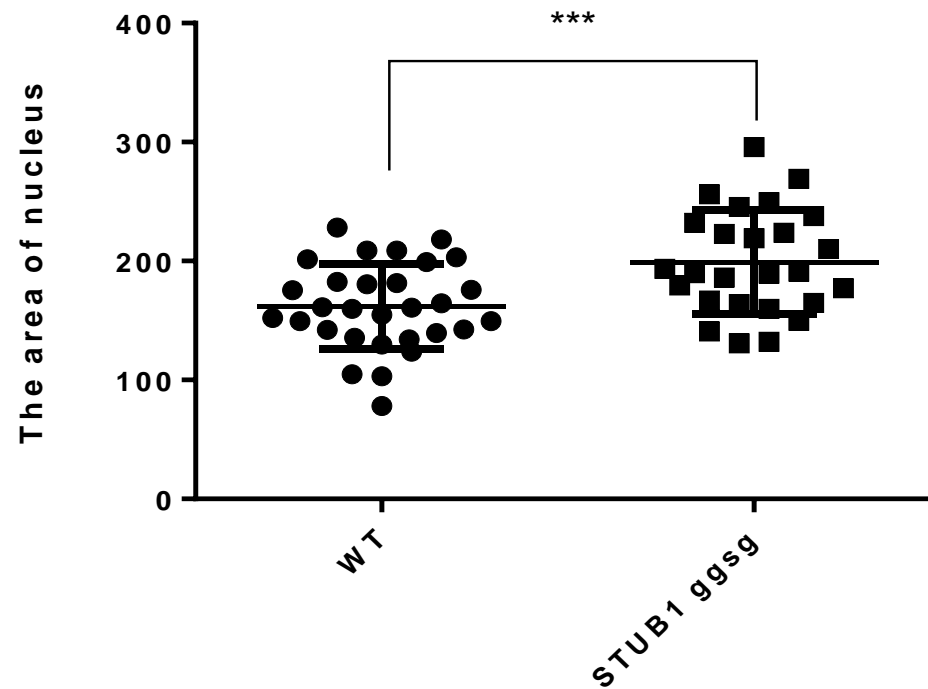
WT



STUB1 ggsg

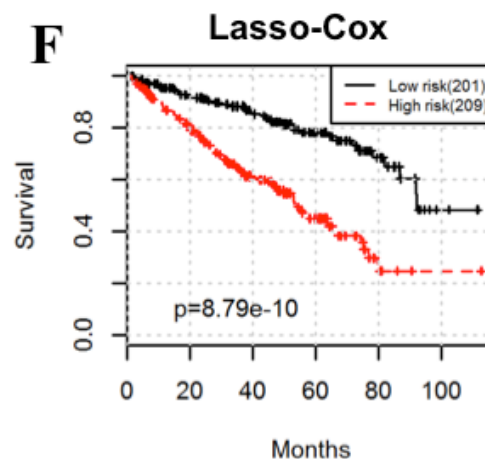
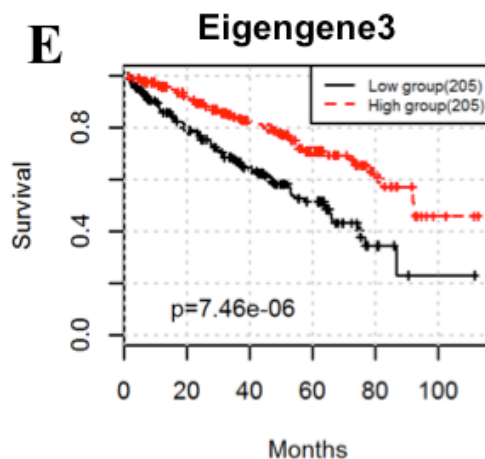
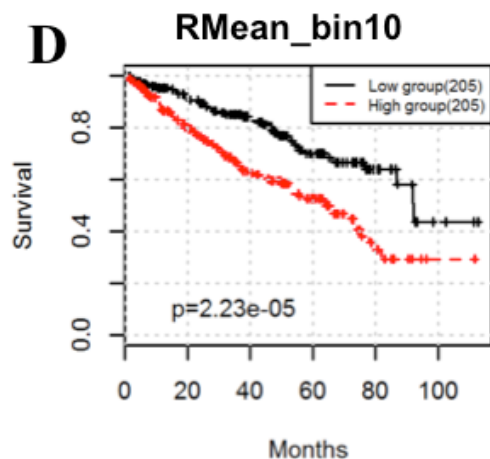
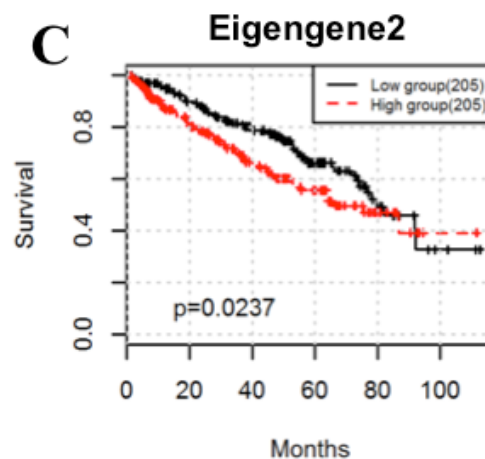
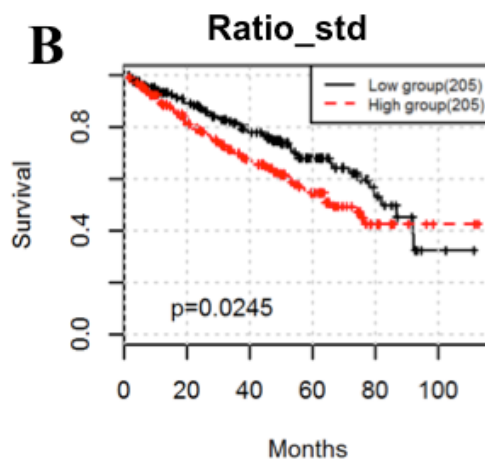
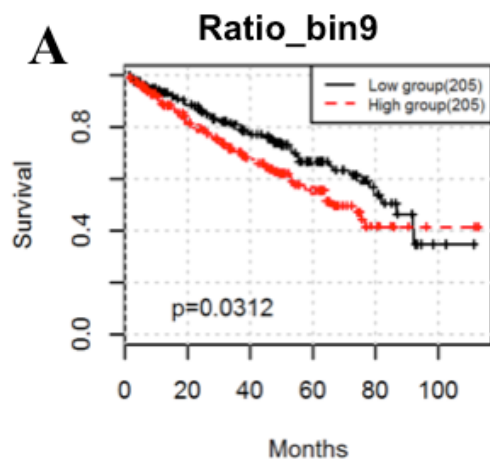


The size of nucleus 48h



Renal Clear Cell Cancer

Question – Can these features do better when combined?

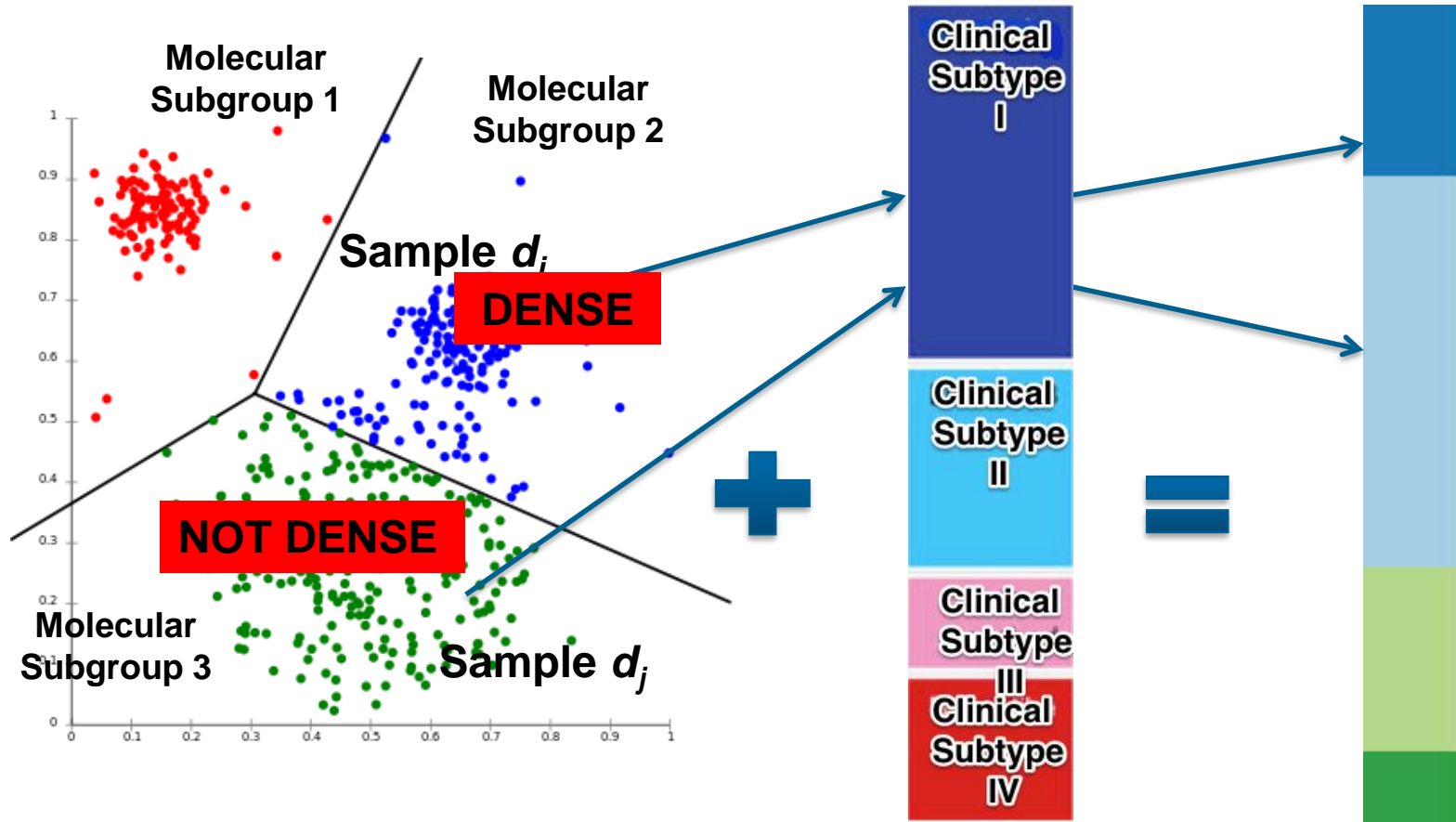


Renal Clear Cell Cancer

Question – Can these features do better when combined?

Variable	Univariate Cox regression		Multivariate Cox regression	
	HR (95% CI)	P value	HR (95% CI)	P value
Lasso-Cox	3.06 (2.10-4.45)	5.02e-9	2.26 (1.46-3.49)	2.31e-4
Clinical				
Grade	2.38 (1.63-3.5)	8.45e-6	1.46 (0.95-2.23)	8.22e-2
Stage	3.68 (2.57-5.27)	1.12e-12	3.00 (2.00-4.49)	9.23e-8
Gene expression				
CSNK2A1	0.90 (0.64-1.26)	5.34e-1	1.07 (0.74-1.56)	7.11e-1
SPP1	1.15 (0.82-1.61)	4.14e-1	1.10 (0.75-1.63)	6.20e-1
DEFB1	1.41 (1.00-1.98)	4.99e-2	1.36 (0.95-1.95)	9.71e-2
PECAM1	0.77 (0.55-1.09)	1.40e-1	1.04 (0.69-1.58)	8.45e-1
EDNRB	0.50 (0.35-0.71)	9.10e-5	0.96 (0.59-1.57)	8.77e-1
TSPAN7	0.54 (0.38-0.76)	5.12e-4	1.03 (0.64-1.67)	9.07e-1
Somatic mutation				
VHL	0.99 (0.70-1.38)	9.33e-1	1.23 (0.86-1.75)	2.57e-1
PBRM1	0.85 (0.58-1.24)	3.94e-1	1.03 (0.69-1.54)	8.85e-1
BAP1	1.49 (0.78-2.85)	2.22e-1	1.49 (0.74-3.00)	2.60e-1
SETD2	1.29 (0.77-2.14)	3.29e-1	1.03 (0.62-1.74)	9.00e-1
TP53	2.26 (1.00-5.15)	5.13e-2	2.86 (1.19-6.86)	1.85e-2

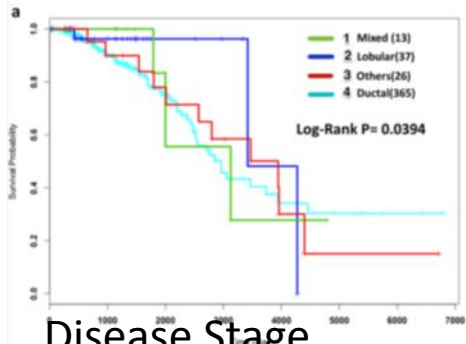
Integration



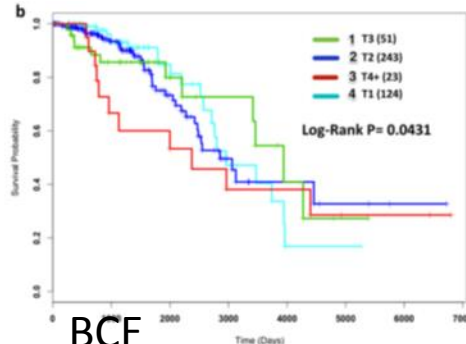
Wang et al, Methods, 2013

Prognosis

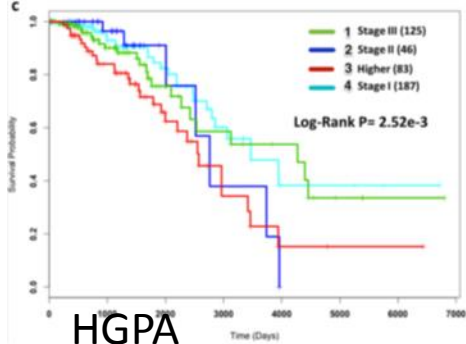
Histology Type



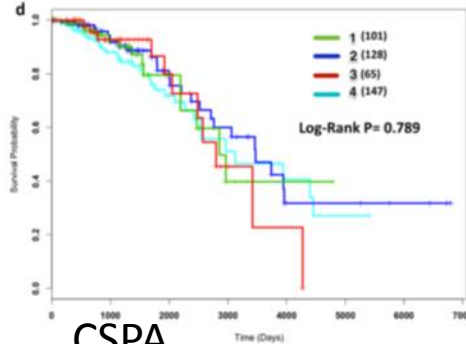
Tumor Grade



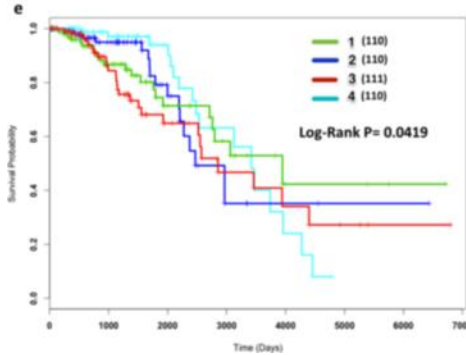
Disease Stage



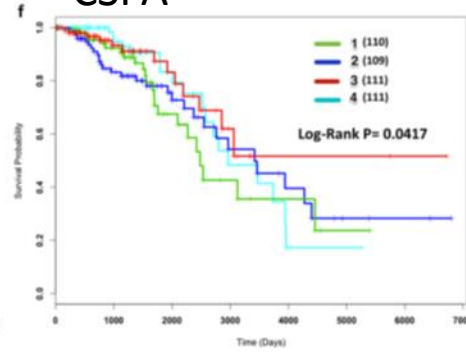
BCE



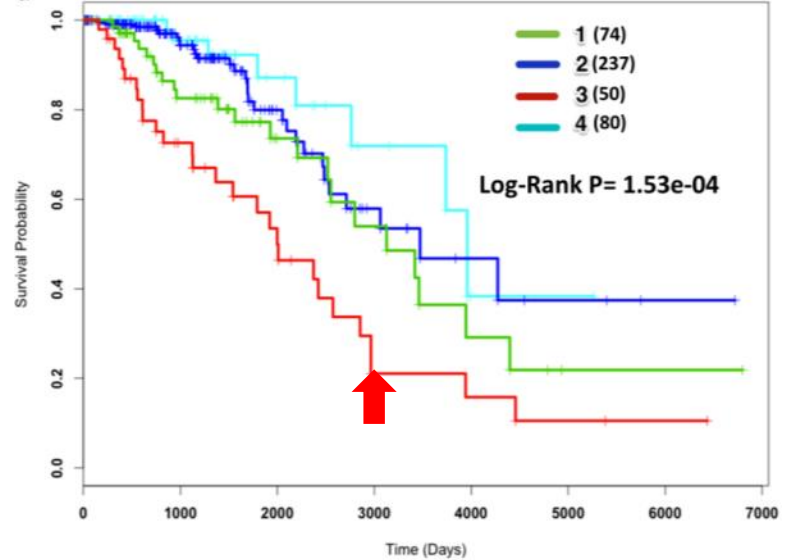
HGPA



CSPA

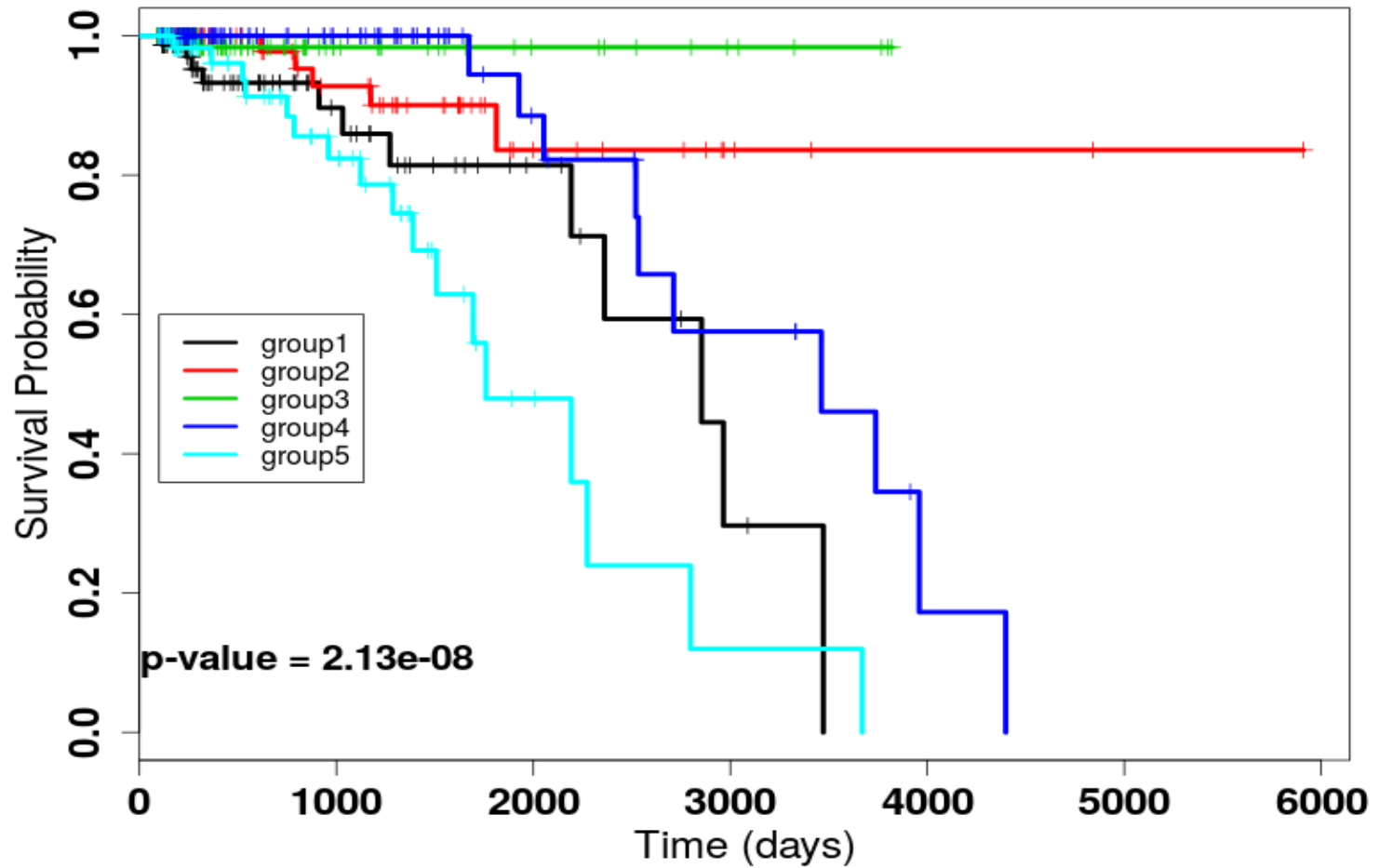


RCP

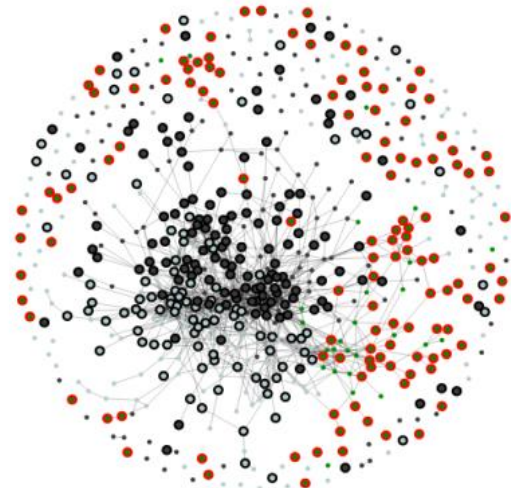
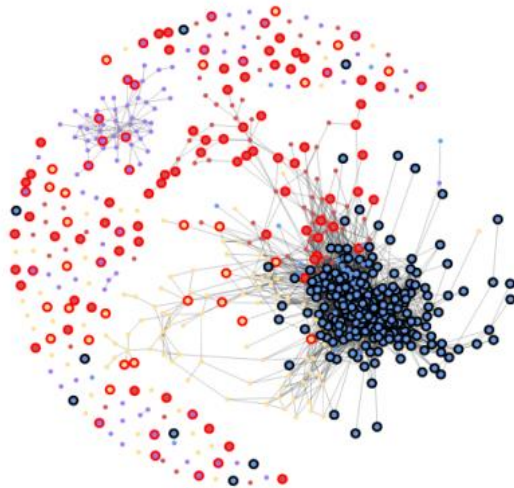
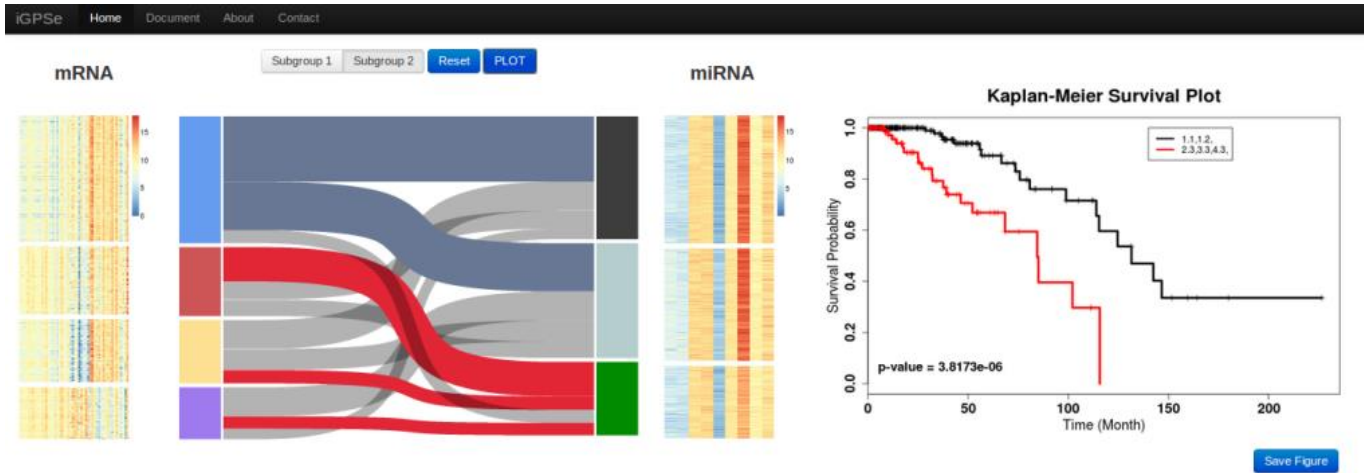


Wang C, Machiraju R, Huang K, *Methods*, 2014

Ongoing – Grassmanian Integration

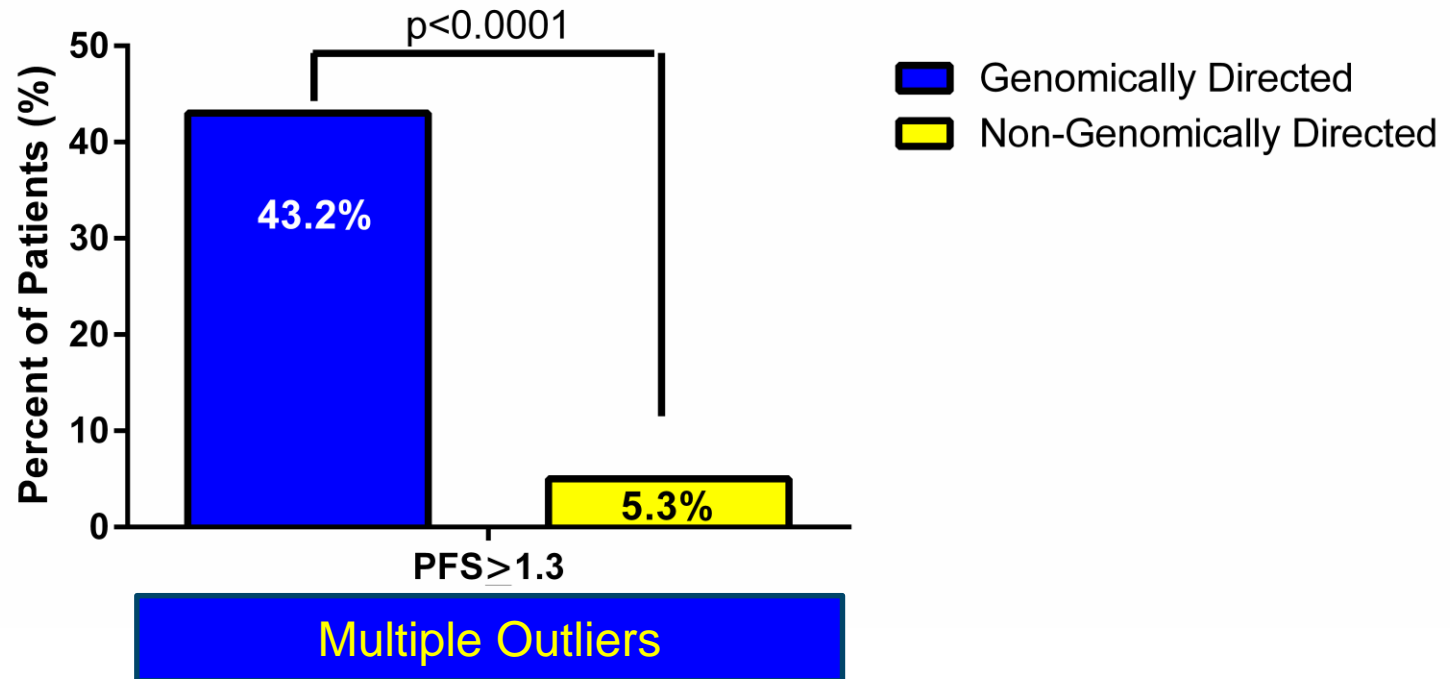


Interactive Patient Stratification



Cancer Precision Medicine Clinic at Indiana University

Proportion of Patients Achieving a PFS Ratio ≥ 1.3



Randovich et al; Oncotarget 2016

Outline

- **Computational Pathology and Precision Medicine**
- **What Information Does Imaging Data Provide - Quantitative Phenotyping Tool**
- **Complementary with Omics Data - Integrative Genomics**
- **Summary and Perspective**

Regarding Data

- Always a lot of data, always not enough data.
- Use public data
 - Not just for "playing", important resource for hypothesis generation
 - For histopathology images
 - TCGA – more than 24,000 large images for cancers (can be visualized at *Cancer Digital Slide Archive*)
 - Allen Brain Atlas – for brain
 - Need to collaborate with domain experts (e.g., pathologists) to use the data meaningfully
 - Find needle in "needle" stack
- Unbalanced data

Summary and Perspective

- **Quantitative phenotyping using imaging**
- **Heterogeneity**
- **Integration of omic and morphological data leads to new biological insights and new ways of patient stratification**
- **Visualization is a useful tool for interactive analysis**
- **High-level phenotype data from EMR**
- **Monitoring and surveillance data, mobile health data**
- **Predictive Marker!**

Acknowledgement

- Chao Wang, Hao Ding, Dr. Shantanu Singh (Broad Institute), Dr. Lee Cooper (Emory U), Dr. Kishore Mosaliganti (HMS), Dr. Sundar Raman, Dr. Hao Ding, Qihang Li, Dr. Liya Ding (HMS), Dr. Jie Zhang, Dr. Nan Meng, Jinchao Di
- Dr. Raghu Machiraju, Dr. Umit Catalyurek, Dr. Jeffrey Parvin, Dr. Parag Mallick (Stanford), Dr. Song-Hai Shi (UIUC), Dr. Lang Li (IUPUI), Dr. Mike Ostrowski, Dr. Lin Yang (IUPUI), Dr. (UKY), Dr. Sashwati Roy, Dr. Joel Saltz (Emory U), Dr. Charles Shapiro, Dr. Debra Zinger, Dr. David Carbone
- Human Frontier Science Program, Nvidia Foundation, NCI ITCR U01, NSF ABI, NCI R01, NCI U54 (ICBP – CCSB), HFSP, HHMI Med-to-Grad Fellowship, Pelotonia Fellowship

Thank you!